
W P S R E H

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 14:32:42 1999; MasPar time 3416.68 Seconds
Tabular output not generated. 1534.145 Million cell updates/sec

Title: >US-08-702-718-1
Description: (1-1891) from US08702718 seq
Perfect Score: 1891
N.A. Sequence: 1 TTTTTCGTTCCATCAGCCTA.....TTATTGATGATATTATGAA 1891
Comp: AAAAAGCAAGGTAGTCGGAT.....ATAAAGTACTATAATAGT

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158
1:em_bal 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pi 14:em_ro 15:em_sts 16:em_vi
Database: genbank11
17:gb_bal 18:gb_ba2 19:gb_hgt1 20:gb_hgt2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pi1
28:gb_pi2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.454; Variance 5.068; scale 2.262
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	1891	100.0	1891	25	A46545 Sequence 1 from Patent	0.00e+00
2	1891	100.0	1891	27	STCIIS S.tuberosum mRNA for m	0.00e+00
3	1238	65.5	1747	25	A46547 Sequence 3 from Patent	0.00e+00
4	1238	65.5	1747	27	N.TCITSYNH N.tabacum mRNA for cit	0.00e+00
5	748	39.6	1859	27	AB017159 Daucus carota mRNA for	0.00e+00
6	700	37.0	1733	27	CMU19481 Citrus maxima citrate	0.00e+00
7	664	35.1	1551	25	A46546 Sequence 2 from Patent	0.00e+00
8	664	35.1	1551	27	BVCITSYNH B.vulgaris mRNA for ci	0.00e+00
9	566	29.8	1680	27	AFMTCITRN Arabidopsis thaliana m	0.00e+00
10	564	29.8	1378	27	PFICITSYNH Populus hybrid mRNA fo	0.00e+00
11	240	12.7	1401	31	AF047042 Homo sapiens citrate s	2.77e-187
12	241	12.7	1455	23	PIGCITSYN Porcine citrate syntha	3.19e-188
13	193	10.2	1481	27	SCMTISNA S.cerevisiae mitochond	1.82e-143

14	193	10.2	2427	27	SCGS01	Yeast gene for citrate	1.82e-143
15	193	10.2	4351	27	SCYNR001C	S.cerevisiae chromsom	1.82e-143
16	193	10.2	2301	27	SCN201952	S.cerevisiae N2019, N2	1.82e-143
17	180	9.5	24643	27	SPAC6C3	S.pombe chromosome 1 c	1.90e-131
18	173	9.1	1800	27	AB001565	Candida tropicalis DNA	5.33e-125
19	138	7.3	5430	27	SCI	Yeast (S.cerevisiae) C	4.69e-93
20	136	7.3	315339	27	SCCHRII	S.cerevisiae chromsom	4.69e-93
21	136	7.2	1680	27	YSCCIT2	Yeast (S.cerevisiae) C	3.01e-91
22	118	6.2	1710	27	ASNCIT1	Aspergillus niger mRNA	4.27e-75
23	100	5.3	2638	27	NEUCIT1A	N.crassa mitochondrial	3.38e-59
24	95	5.0	562	34	KLAJ9626	Kluyveromyces lactis D	7.74e-55
25	95	5.0	1497	21	TETCITSYN	T. thermophila mRNA fo	7.74e-55
26	84	4.4	48096	21	CET20G5	Caenorhabditis elegans	2.34e-45
27	79	4.2	118327	28	ATAC004521	Arabidopsis thaliana c	4.18e-41
28	73	3.9	152184	19	CEV70G10	Caenorhabditis elegans	4.64e-36
29	67	3.5	467	34	GL3212	human STS WI-12459.	4.36e-31
30	61	3.2	2790	27	SCCI13SYN	S.cerevisiae citrate s	3.36e-26
31	61	3.2	17660	27	SC9723	S.cerevisiae chromsom	3.36e-26
32	61	3.2	38779	27	SCU31900	Saccharomyces cerevisi	3.36e-26
33	61	3.2	165536	27	SCCHRXV1	S.cerevisiae chromsom	3.36e-26
34	52	2.7	2398	28	ENU89675	Emericella nidulans ci	4.53e-19
35	45	2.4	7218	25	I66494	Sequence 14 from paten	9.66e-14
36	42	2.2	7218	25	I66494	Sequence 14 from paten	1.57e-11
37	40	2.1	74371	31	AC005369	Homo sapiens chromsom	4.36e-10
38	38	2.0	1056	23	MVU87256	Mustela vison GT dinuc	1.14e-08
39	35	1.9	1056	23	MVU87256	Mustela vison GT dinuc	1.34e-06
40	34	1.8	965	25	AR024229	Sequence 22 from paten	6.35e-06
41	33	1.7	10772	21	AF012089	Drosophila melanogaste	2.94e-05
42	32	1.7	10772	21	AF012089	Drosophila melanogaste	1.33e-04
43	31	1.6	215	25	I28278	Sequence 5 from patent	5.87e-04
44	31	1.6	215	25	I28278	Sequence 5 from patent	5.87e-04
45	30	1.6	122247	19	AC006419	Arabidopsis thaliana ,	2.53e-03

ALIGNMENTS

RESULT	1	A46545	1891 bp	DNA	PAT	07-MAR-1997
LOCUS		Sequence 1 from Patent	WO9524487.			
DEFINITION		A46545				
ACCESSION		G2300707				
NID		A46545.1	GI:2300707			
VERSION						
KEYWORDS		potato.				
SOURCE		Solanum tuberosum				
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta, eudicotyledons, Asteridae; Solanales; Solanales; Solanales; Solanales; Solanales; section Petota.				
REFERENCE		1 (bases 1 to 1891)				
AUTHORS		Mueller-Roeber, B., Landschuetz, V. and La, C. U.				
TITLE		PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN PLANTS				
JOURNAL		Patent: WO 9524487-A 1 14-SEP-1995;				
COMMENT		HOECHST SCHERING AGREVO GMBH (DE) Other publication DE 4438821 960425 Other publication DE 4435366 960328 Other publication AU 2067995 950925 Other publication DE 4408629 950914.				

FEATURES	Location/Qualifiers
source	1..1891
	/organism="Solanum tuberosum"
	/strain="C.V. DESIREE"
	/db_xref="taxon:4113"
	/tissue_type="LEAF"
	/clone_lib="CDNA LIBRARY IN PBLUESCRIPTKS"
	/clone="PCBS"
	73..1486
	/EC_number="4.1.3.7"
	/codon_start=1
	/product="CITRATE SYNTHASE"
	/protein_id="CAA02908.1"
	/db_xref="PID:e308252"
CDS	

```

/db_xref="PID:q2300708"
/db_xref="GI:2300708"
/translation="MFVRSVSLSKLRSRAVOOSNVSNRVLQVOTSSGLDLRSEL
VQELIPEQDRLKKTKSDMKSGIGNITVDMVLGMPGMTGLMKPHYLDPEGIRPRG
LSIPBCQVLPAAKPGVPLPEGLLWLLTKVPSKEQVNSIVSGHAEISLIIMY
TTIDALPVTAHPMTQFATGVMAOVQSEFOKXKHKYWEPTVEDSMNLIAOV
LVAAYVRYMKNGDIPKDESLOYGANFAMHGLFSSSEMHELLMLRYTHIHSDESG
NYSATGHVLASALSDPYLSPFAALNGLAGPLHGLAQEVLLWIKSVBECGNISKE
QUKDYVHLNSGVKVPFGHGVLRKTVPRYTQREFAMKHPEDPLFQLVSKLYEF
LLFLONLAKLPWNPDAHSGVLLNYGLTEARYTYTLFVGSRALGICSLIWDRAVG
LPLERPKSVTMEWLENQCKA"
BASE COUNT      512 a   370 c   425 g   584 t
ORIGIN

```

```

Query Match      100.0%; Score 1891; DB 25; Length 1891;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TTTTTCGTTCCATCAGCCTACTTGAGATGATATTCACCTGGTAAAGATTAAATTTTTTGA 60
QY 1 TTTTTCGTTCCATCAGCCTACTTGAGATGATATTCACCTGGTAAAGATTAAATTTTTTGA 60

Db 61 TTTTCGGAGCAATGGTGTACCGTAGCGTTTCGTTGCTGTCAAAGCTCCGCTCTCGA 120
QY 61 TTTTCGGAGCAATGGTGTACCGTAGCGTTTCGTTGCTGTCAAAGCTCCGCTCTCGA 120

Db 121 GCGGTCCACAGTCAATGTTAGCAATCTGTGCGCTGGCTCAAGTCCAAACCTCTCC 180
QY 121 GCGGTCCACAGTCAATGTTAGCAATCTGTGCGCTGGCTCAAGTCCAAACCTCTCC 180

Db 181 GGTCTTGATCGGTTCTCGAGTGGTACAAAGATTGATTCCTGAAACAGAGATCGCGTG 240
QY 181 GGTCTTGATCGGTTCTCGAGTGGTACAAAGATTGATTCCTGAAACAGAGATCGCGTG 240

Db 241 AAAAGATCAAGTCAATGAAAGTTCAATTTGGACATCACAGTTGATATGGTCTT 300
QY 241 AAAAGATCAAGTCAATGAAAGTTCAATTTGGACATCACAGTTGATATGGTCTT 300

Db 301 GGTGGAATGAGGAATGACAGATTACTGTGGAACCTCATTAACCTTGACCCCTGATGAG 360
QY 301 GGTGGAATGAGGAATGACAGATTACTGTGGAACCTCATTAACCTTGACCCCTGATGAG 360

Db 361 GGAATTGCTTCGGGGTGTCTATACCTGAAATCCCAAAAGGATTAACCTGCAGCAAG 420
QY 361 GGAATTGCTTCGGGGTGTCTATACCTGAAATCCCAAAAGGATTAACCTGCAGCAAG 420

Db 421 CCTGGGGTTCAGCGCTTCGCTGAGTCTCTCTGGTCTCTTTTAAACAGAAAGTGCCA 480
QY 421 CCTGGGGTTCAGCGCTTCGCTGAGTCTCTCTGGTCTCTTTTAAACAGAAAGTGCCA 480

Db 481 TCAAAAGAGCAAGTGAATTCAATTGTCTCAGGAATTGACAGATCGGGGATCATATCCCTG 540
QY 481 TCAAAAGAGCAAGTGAATTCAATTGTCTCAGGAATTGACAGATCGGGGATCATATCCCTG 540

Db 541 ATCATATGATACAACTATGATGCTTACCAGTACAGCTCATCCAAATGACCCAGTTT 600
QY 541 ATCATATGATACAACTATGATGCTTACCAGTACAGCTCATCCAAATGACCCAGTTT 600

Db 601 GCTACTGGAGTATGCTCTTCAGGTTCAAAGTGAATTTCAAAGGCAATACGAGAAAGGG 660
QY 601 GCTACTGGAGTATGCTCTTCAGGTTCAAAGTGAATTTCAAAGGCAATACGAGAAAGGG 660

Db 661 ATTCACAAATCAAAGTATTGGGAACCAACATATGAGGATTCATGAATCTGATTGCTCAA 720
QY 661 ATTCACAAATCAAAGTATTGGGAACCAACATATGAGGATTCATGAATCTGATTGCTCAA 720

Db 721 GTTCCACTTGCTGCTTATGTTATCCAGGATGTACAGAAATGGTGACACTATACCT 780
QY 721 GTTCCACTTGCTGCTTATGTTATCCAGGATGTACAGAAATGGTGACACTATACCT 780

Db 781 AAGGATGAATCCCTGATTATGGTCAAAATTTTGGTCAATGTTGTTTTCAGTAGTCT 840
QY 781 AAGGATGAATCCCTGATTATGGTCAAAATTTTGGTCAATGTTGTTTTCAGTAGTCT 840

```

```

Db 841 GAAATGCATGAACCTCTCTATGAGGCTCTATATAACAATACACAGTATGATCAAGGTGGT 900
QY 841 GAAATGCATGAACCTCTCTATGAGGCTCTATATAACAATACACAGTATGATCAAGGTGGT 900

Db 901 AATGTCAGTGTCTACACCGGTCACTTGGTCTGCTAGTCTTGTCTGATCCTTACCTCTCC 960
QY 901 AATGTCAGTGTCTACACCGGTCACTTGGTCTGCTAGTCTTGTCTGATCCTTACCTCTCC 960

Db 961 TTTGCTGCTGCTTTTGAATGGTTTAGCCGGACCACTTTCATGGTTTAGCCAATCAGGAAGTT 1020
QY 961 TTTGCTGCTGCTTTTGAATGGTTTAGCCGGACCACTTTCATGGTTTAGCCAATCAGGAAGTT 1020

Db 1021 TTGCTATGATAAATCTGTTGTAGAGAATGTGGGAGAACATTTCCAAAGAGCAGTTG 1080
QY 1021 TTGCTATGATAAATCTGTTGTAGAGAATGTGGGAGAACATTTCCAAAGAGCAGTTG 1080

Db 1081 AAAGACTATGTTTGAACCAATTTGAACAGTGGCAAGTGTTCCTGTTTGGACATGGA 1140
QY 1081 AAAGACTATGTTTGAACCAATTTGAACAGTGGCAAGTGTTCCTGTTTGGACATGGA 1140

Db 1141 GTTCTGCGAAGACTGTACCAAGATATACATGCCAGAGAGAGTTTCGCTATCAAGCATTG 1200
QY 1141 GTTCTGCGAAGACTGTACCAAGATATACATGCCAGAGAGAGTTTCGCTATCAAGCATTG 1200

Db 1201 CCGTGAAGATCCACTGTTTCAACTGGTTTCAAACTCTACGAAGTTTTCCTCTCTCTTA 1260
QY 1201 CCGTGAAGATCCACTGTTTCAACTGGTTTCAAACTCTACGAAGTTTTCCTCTCTCTTA 1260

Db 1261 CAGAACTTTGGCAAGTTTAAACCTTTGGCCAAATTTGATGCCACAGTGTGTTGTTG 1320
QY 1261 CAGAACTTTGGCAAGTTTAAACCTTTGGCCAAATTTGATGCCACAGTGTGTTGTTG 1320

Db 1321 AACTATATGTTTAACTGAAGCAAGATATATACGGTCTCTTTTGGTATTAAGAACT 1380
QY 1321 AACTATATGTTTAACTGAAGCAAGATATATACGGTCTCTTTTGGTATTAAGAACT 1380

Db 1381 CTTGGCATTTGCTCTCAGCTAATTTGGGACCGAGCTCTTGGAATTCGGCTAGAGAGGCCA 1440
QY 1381 CTTGGCATTTGCTCTCAGCTAATTTGGGACCGAGCTCTTGGAATTCGGCTAGAGAGGCCA 1440

Db 1441 AAGAGTGTCAATGGAGTGTCTTGAGAACCCAGTGCAGAACGATGAATTTGTTGAAAT 1500
QY 1441 AAGAGTGTCAATGGAGTGTCTTGAGAACCCAGTGCAGAACGATGAATTTGTTGAAAT 1500

Db 1501 CTCGGGAGCATAAACACAATGTATATCTCTATGAATTAATTTGCTTACAAAGCAGTCTCT 1560
QY 1501 CTCGGGAGCATAAACACAATGTATATCTCTATGAATTAATTTGCTTACAAAGCAGTCTCT 1560

Db 1561 TTCTTTGGGGGACAAGATAGGTTCGGCCCTTCAATGGGTTAAAGAACTTCAGTTCAAACTTC 1620
QY 1561 TTCTTTGGGGGACAAGATAGGTTCGGCCCTTCAATGGGTTAAAGAACTTCAGTTCAAACTTC 1620

Db 1621 ACTGAATTTGTTGTAATTTGATGTTTCTCGAGACTTGTCTGAAATTTTGAACCTTAGTCT 1680
QY 1621 ACTGAATTTGTTGTAATTTGATGTTTCTCGAGACTTGTCTGAAATTTTGAACCTTAGTCT 1680

Db 1681 AGTGAATTCATTTTCTTCATTTCGGAATTTCTCACAGCTGATCCAGCAGTGTAAATAATTA 1740
QY 1681 AGTGAATTCATTTTCTTCATTTCGGAATTTCTCACAGCTGATCCAGCAGTGTAAATAATTA 1740

Db 1741 ATAGTCAATGCTATTAATTCGCTTCTTGGTTGCCATTAGACTTTTGAATGACTTCTT 1800
QY 1741 ATAGTCAATGCTATTAATTCGCTTCTTGGTTGCCATTAGACTTTTGAATGACTTCTT 1800

Db 1801 GCTGGAAGTGTAGTAATCGGCTGATTACCGCAATAAAGTCAATTTGTAGTTCTTAAA 1860
QY 1801 GCTGGAAGTGTAGTAATCGGCTGATTACCGCAATAAAGTCAATTTGTAGTTCTTAAA 1860

Db 1861 TTTGCTAATCTCTATTTGATGATATATGAA 1891
QY 1861 TTTGCTAATCTCTATTTGATGATATATGAA 1891

```



```

RESULT 2
LOCUS 1891 bp mRNA PLN 20-MAR-1996
DEFINITION S.tuberosum mRNA for mitochondrial citrate-synthase.
ACCESSION X75082
NID 9483509
VERSION X75082.1 GI:483509
KEYWORDS mitochondrial citrate synthase.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanales; Solanaceae; Solanum; Potatoe;
section Petota.
1 (bases 1 to 1891)
Landshuetze.V.
Direct Submission
Submitted (20-SEP-1993) V. Landshuetze, Inst. fuer Genbiologische
Forschung, Berlin GmbH, Ihnestr. 63, 14195 Berlin, FRG
2 (bases 1 to 1891)
Landshuetze.V., Muller-Rober,B. and Willmitzer,L.
Mitochondrial citrate synthase from potato: predominant expression
in mature leaves and young flower buds
Planta 196 (4), 756-764 (1995)
96073431
Location/Qualifiers
1. 1891
/organism="Solanum tuberosum"
/cultivar="Desiree"
/db_xref="taxon:4113"
/dev_stage="mature"
/clone_lib="leaf library"
/clone="5-25-10"
73. 1488
/EC_number="4.3.1.7"
/note="ethanolamine ammonia-lyase"
/codon_start=1
/product="mitochondrial citrate-synthase"
/protein_id="CAA52976.1"
/db_xref="PID:483510"
/db_xref="GI:483510"
/db_xref="SPTREMBL:Q43175"
/translation="MFYRSVLSKLSRAVQSNVSNVRVQVOTSSGLDLRSEL
VQVLEPEQDRUKKIKSDMKSGISNITVDMVLGMRGTGLLWKPHYLDPEGIFPFG
LSIQKQVLPAAKPGPELPEGLLWLLITGKVPKSEQVNSIVSGIAESILIIIMY
TIDALPTVAHPMTQFATGMALQVSEFOKAYEKGHKSKEWPTVECSNMLIAQVP
LVAAVYRMYKNGDTIPKDESLDYGANFAHMLGFSSEMHELMLRYVTHSDHEGS
NVAHTGHLVASALSDPYLSFAALNGLAGPLHGLANOELLWIKSVVEGGENISKE
QLKDYVWKLNSGKVPFGHGVLRKTPRTVCQREFAKMLPEDPLFQLVSKLYEF
LFLQNLAKLKPWPNVDHSGVLLNYGLTEARYTYVLFGVSALGICSLINDRAIG
LPLERPKSVTMEWLENQCKKA"
BASE COUNT 512 a 370 c 425 g 584 t
ORIGIN
Query Match 100.0%; Score 1891; DB 27; Length 1891;
Best Local Similarity 100.0%; Pred. No. 0.0ne-00;
Matches 1891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 TTTTTCGTCATCAGCCTACTTGAGATGATTCCACACTGGTAAAGTAAATTTTTTGA 60
Qy 1 TTTTTCGTCATCAGCCTACTTGAGATGATTCCACACTGGTAAAGTAAATTTTTTGA 60
Db 61 TTTTCGGCAGCAATGGTGTCTACCGTAGCGTTTCGTTGCTGTCAAGTCGCGCTCGA 120
Qy 61 TTTTCGGCAGCAATGGTGTCTACCGTAGCGTTTCGTTGCTGTCAAGTCGCGCTCGA 120
Db 121 CGCGTCCAACAGTCAATGTTAGCAATTCCTGCGCTGGCTTCAAGTCCAAACCTCTCC 180
Qy 121 CGCGTCCAACAGTCAATGTTAGCAATTCCTGCGCTGGCTTCAAGTCCAAACCTCTCC 180
Db 181 GGTCTTGATCTCGCTTCTGAGCTGGTACAAAGATTGATTCCTGAACAACAGGATCGCCTG 240
Qy 181 GGTCTTGATCTCGCTTCTGAGCTGGTACAAAGATTGATTCCTGAACAACAGGATCGCCTG 240

```

```

Db 241 AAAAATCAAGTCAGATATGAAGGTTCAATTGGGAACATCACAGTTGATGTTCTT 300
Qy 241 AAAAATCAAGTCAGATATGAAGGTTCAATTGGGAACATCACAGTTGATGTTCTT 300
Db 301 GGTGAATGAGAGGAATGACAGGATTACTGTGGAACCTCATTTACCTTGACCTGAG 360
Qy 301 GGTGAATGAGAGGAATGACAGGATTACTGTGGAACCTCATTTACCTTGACCTGAG 360
Db 361 GGAATTCGCTTCGGGGGTTGCTATACCTGAATCCAAAAGGTTATTACCTTGACCAAG 420
Qy 361 GGAATTCGCTTCGGGGGTTGCTATACCTGAATCCAAAAGGTTATTACCTTGACCAAG 420
Db 421 CTTGGGGTGAGCCCTTGCTGCTGAAGGTTCTCTGGCTCTTTTAAACAGAAAGTGCCA 480
Qy 421 CTTGGGGTGAGCCCTTGCTGCTGAAGGTTCTCTGGCTCTTTTAAACAGAAAGTGCCA 480
Db 481 TCAAAAGACAAGTGAATTCATTTCTCAGGAATTGAGAGTCGGGCATCATATCCCTG 540
Qy 481 TCAAAAGACAAGTGAATTCATTTCTCAGGAATTGAGAGTCGGGCATCATATCCCTG 540
Db 541 ATCATCATGTATACAACTATTGATGCCITACCAGTCACAGCTCATCCAATGACCCAGTT 600
Qy 541 ATCATCATGTATACAACTATTGATGCCITACCAGTCACAGCTCATCCAATGACCCAGTT 600
Db 601 GCTACTGGAGTCATGGCTCTTCAGGTTCAAAGTGAATTTCAAAGGTCATACGAGAAAGGG 660
Qy 601 GCTACTGGAGTCATGGCTCTTCAGGTTCAAAGTGAATTTCAAAGGTCATACGAGAAAGGG 660
Db 661 ATTCAAAATCAAAGTATTGGGAACCAACATATGAGGATTCGATGAATCTGATGCTCAA 720
Qy 661 ATTCAAAATCAAAGTATTGGGAACCAACATATGAGGATTCGATGAATCTGATGCTCAA 720
Db 721 GTTCCACTGTTGCTGCTTATGTTATCGCAGATGTACAAGATGGTGACACTATACCT 780
Qy 721 GTTCCACTGTTGCTGCTTATGTTATCGCAGATGTACAAGATGGTGACACTATACCT 780
Db 781 AAGGATGAATCCCTGGATTATGGTCAAAATTTTGCTCACATGCTTGTTTTCAGTAGCTCT 840
Qy 781 AAGGATGAATCCCTGGATTATGGTCAAAATTTTGCTCACATGCTTGTTTTCAGTAGCTCT 840
Db 841 GAAATGCATGAAGTCTTATGAGGCTCTATGTAACAAATACACAGTGATCATGAAGGTGGT 900
Qy 841 GAAATGCATGAAGTCTTATGAGGCTCTATGTAACAAATACACAGTGATCATGAAGGTGGT 900
Db 901 AATGTCAGTGTCTCACACCGGTCTACTTGGTGTAGTGTCTGATGCTTACCTCTCC 960
Qy 901 AATGTCAGTGTCTCACACCGGTCTACTTGGTGTAGTGTCTGATGCTTACCTCTCC 960
Db 961 TTTGCTGCTGTTTGAATGGTTTAGCCGACACCTTCATGTTTGAAGTAAATGAGGAGTT 1020
Qy 961 TTTGCTGCTGTTTGAATGGTTTAGCCGACACCTTCATGTTTGAAGTAAATGAGGAGTT 1020
Db 1021 TTGCTATGATTAATCTGTTAGAGGATGTGGGAGAACATTTCCAAAGACAGTTG 1080
Qy 1021 TTGCTATGATTAATCTGTTAGAGGATGTGGGAGAACATTTCCAAAGACAGTTG 1080
Db 1081 AAAGACTATGTTTGAAGAACATTTGAACAGTGGCAAGGTTGTCCTGGTTTGGACATGGA 1140
Qy 1081 AAAGACTATGTTTGAAGAACATTTGAACAGTGGCAAGGTTGTCCTGGTTTGGACATGGA 1140
Db 1141 GTTCTCGGAAGACTGTACCAAGATATACATGCCAGAGAGTTCCGCTATGAAGCAATTG 1200
Qy 1141 GTTCTCGGAAGACTGTACCAAGATATACATGCCAGAGAGTTCCGCTATGAAGCAATTG 1200
Db 1201 CTTGAAGATCCACTGTTTCAACTGTTTCAAACTCTACGAAGTTTTCCTCTGTTCTTA 1260
Qy 1201 CTTGAAGATCCACTGTTTCAACTGTTTCAAACTCTACGAAGTTTTCCTCTGTTCTTA 1260
Db 1261 CAGAAGCTTGGCAAGTAAAGCTTGGCCAAATGTTGATGCCACAGTGGTGTGTTG 1320
Qy 1261 CAGAAGCTTGGCAAGTAAAGCTTGGCCAAATGTTGATGCCACAGTGGTGTGTTG 1320

```

```

Db 1321 AACATTATGGTTTAACTGAACAGATATATATACGGTCCCTTTGGCGTATCAAGAGCT 1380
QY 1321 AACATTATGGTTTAACTGAACAGATATATATACGGTCCCTTTGGCGTATCAAGAGCT 1380
Db 1381 CTTGGCATTGTCTCAGCTAATTTGGGACCGAGCTCTTGGATTGGCGCTAGAGAGGCCA 1440
QY 1381 CTTGGCATTGTCTCAGCTAATTTGGGACCGAGCTCTTGGATTGGCGCTAGAGAGGCCA 1440
Db 1441 AGAGTGTCAATAGAGTGGCTTGAGAACAGCTGCAAGAAAGCATGAATTTGTTGAAAT 1500
QY 1441 AGAGTGTCAATAGAGTGGCTTGAGAACAGCTGCAAGAAAGCATGAATTTGTTGAAAT 1500
Db 1501 CTGCGGAGTAAACACAACTATATCTCTATGAATAATTTGCTGACAAAGCACTCCT 1560
QY 1501 CTGCGGAGTAAACACAACTATATCTCTATGAATAATTTGCTGACAAAGCACTCCT 1560
Db 1561 TTCTTGGGACAAAGATAGTGGCGCCCTTCAATGGGTTTAAAGCACTTCAAGTCAAACTTC 1620
QY 1561 TTCTTGGGACAAAGATAGTGGCGCCCTTCAATGGGTTTAAAGCACTTCAAGTCAAACTTC 1620
Db 1621 ACTGAATTTGTGTAATTTGTATGGTTTCTCGAGAGCTTGTCTGTAATTTGAACTTAGTCT 1680
QY 1621 ACTGAATTTGTGTAATTTGTATGGTTTCTCGAGAGCTTGTCTGTAATTTGAACTTAGTCT 1680
Db 1681 AGTGAATTCATTTTCTTCAATCGGAATTCCTCACAGCTGATCGAGCATGTAAAAATTA 1740
QY 1681 AGTGAATTCATTTTCTTCAATCGGAATTCCTCACAGCTGATCGAGCATGTAAAAATTA 1740
Db 1741 ATAGGTCAATAGTATTAATTCGCTGCTTCTTGGTGGCATTAGACTTGTGAATTCCTTT 1800
QY 1741 ATAGGTCAATAGTATTAATTCGCTGCTTCTTGGTGGCATTAGACTTGTGAATTCCTTT 1800
Db 1801 GCTGGAAGTGTAGTAACTGGCTGATTCACGCAATAAAGCTGCAATTTGTCTTTTAA 1860
QY 1801 GCTGGAAGTGTAGTAACTGGCTGATTCACGCAATAAAGCTGCAATTTGTCTTTTAA 1860
Db 1861 TTGTCTAATTTCTATTGATGATATTGAA 1891
QY 1861 TTGTCTAATTTCTATTGATGATATTGAA 1891

RESULT 3
LOCUS A46547 1747 bp DNA PAT 07-MAP-1997
DEFINITION Sequence 3 from Patent WO9524487.
ACCESSION A46547
NID 92300711
VERSION A46547.1 GI:2300711
KEYWORDS common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanaceae; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1747)
AUTHORS Mueller-Roeber,B., Landschuetz,V. and La.C.U
TITLE PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN
PLANTS
JOURNAL Patent: WO 9524487-A 3 14-SEP-1995;
HOECHST SCHERING AGREVO GMBH (DE)
COMMENT Other publication DE 4438821 960425
Other publication DE 4435368 960328
Other publication AU 2067995 950925
Other publication DE 4408629 950914.
FEATURES
Location/Qualifiers
1. 1747
/organism="Nicotiana tabacum"
/db_xref="taxon:4097"
/tissue_type="LEAF"
/clone="TCS"
70 1479
/EC_number="4.1.3.7"
/codon_start=1
/product="CITRATE SYNTHASE"
CDS

```

```

/protein_id="CAA02910.1"
/db_xref="PID:e306253"
/db_xref="PID:g2300712"
/db_xref="GI:2300712"
/translation="MVFYRGVSLLSKLRSRAVQOITNLSNVRWLOVOTSSGLDLRSEL
QELIPEQDDPLKKLSEHGKVLGNITVDMVLGGMPMTGLLWETSLIDPDSIFRPG
LSIYECOKVLPAAKPGGEPLPEGLMLLLTGKVPKSEVDLSLSOELRSRATVPDVYK
TIDALPYATPMTQFAIGVWALQVQSEFAKKEGHIKSWEPYIEISMSLIAVPL
VAAYVYRMKYKNGNTIPKDDSLDYGANFHMILGFSDDMHMLKLYITVTHSDEGNNV
SDHTGHLVASALSDPYLSFAAALINGLAPLHGLANOELLWKISVVEEGENISKEQL
KDAWMTKIKSGVPGFGRKRTDPRYTQREFALKHPEDPFLQVAKLYEVLQ
FLONLAKNPWNPVDAHSGVLLNYGLTEARYTVLVFGVSRALGICSQLINDRALGLP
LERPKSVTMEMLNHCKKA"
BASE COUNT 490 a 335 c 400 g 522 t
ORIGIN
Query Match 65.5%; Score 1238; DB 25; Length 1747;
Best Local Similarity 92.8%; Pred. No. 0.00e+00;
Matches 1439; Conservative 0; Mismatches 93; Indels 18; Gaps 14;
Db 38 GGTAAAGTTAATTGTT-GAATTTTGGAGCCATGGTGTCTTATCCGCGCTTCTCTG 96
QY 40 GGTAAAGTTAATTTTGTGATTTTCGCGAGCAATGGTGTCTTACCGTAGCGTTCTG 99
Db 97 CTGTCAAAGCTCGCTTCGAGCGGTCCACAGACAAATCTTAGCAACTCTGTCCGCTG 156
QY 100 CTGTCAAAGCTCGCTTCGAGCGGTCCACAGCAATCTTAGCAACTCTGTCCGCTG 159
Db 157 CTTCAGTCCAAAGCTCTTCTGCTTGTGATCTGCTTCTGAGCTG---CAAGAAATGAT 213
QY 160 CTTCAGTCCAAAGCTCTTCTGCTTGTGATCTGCTTCTGAGCTGTTACAGAAATGAT 219
Db 214 CCAGACACAGAGTCCGCTAAAGCAAGCTCAAGTCAGAGTATGSAAGGTTCAATTTGG 273
QY 220 CCTGAACACAGAGTCCGCTGAAAGAGTCAAGTCAGAT-ATG-AAAGGTTCAATTTGG 277
Db 274 AACATCACAGTTGATATGTTCTTGGTGGATGAGAGAAATGACAGGATTTACTTGGGA 333
QY 278 A-CATCACAGTTGATATGTTCTTGGTGGATGAGAGAAATGACAGGATTTACTTGG-AA 335
Db 334 ACCTCATTAC-TTGACCCCGATGAAGAAATTCGCTTTTCGGGGCTTGTCTATCTATGAAT 392
QY 336 ACCTCATTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 395
Db 393 CCAAAAGTATTACCTGAGCAAGAGCTTGGGGAGAGCCCTTGGCTGAAGGTTCTTCTG 452
QY 396 CCAAAAGTATTACCTGAGCAAGAGCTTGGGGAGAGCCCTTGGCTGAAGGTTCTTCTG 455
Db 453 GCTTCTTTTAAAGAGAGAGTGGCCATCAAAAGAGCAAGTGGATTCA-TTGTCTCAGGAAT 511
QY 456 GCTTCTTTTAAAGAGAGAGTGGCCATCAAAAGAGCAAGTGGATTCAATTTCTCAGGAAT 515
Db 512 TCGGAAGTCTGTGTAC-TGTCCCGATCATG-TATACAAAAGTATTGATGCTTACAGT 569
QY 516 TCGAGAGTGGGGATCATATCCCTGATCATCATGATATACAACTATTGATGCTTACAGT 575
Db 570 CACAGCTCATCCAACTGACTGACTTGTGCTAGTGAATCATGCTTTCAGGTTTCAAGATCA 629
QY 576 CACAGCTCATCCAACTGACTGACTTGTGCTAGTGAATCATGCTTTCAGGTTTCAAGTCA 635
Db 630 ATTCAAAGGATATGAGAAAGGATTCACAAATCAAGTTATGGGACCGACATATCA 689
QY 636 ATTTCAAAGGATATGAGAAAGGATTCACAAATCAAGTTATGGGACCGACATATCA 695
Db 690 GGATTCAGATGTTGATTGCTCAAGTTCAGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 749
QY 696 GGATTCAGATGTTGATTGCTCAAGTTCAGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 755
Db 750 GTACAAGAGCGCAACACTATACCTAAGGATGACTCACTGATGATGATGATGATGATGATG 809
QY 756 GTACAAGAGCGCAACACTATACCTAAGGATGACTCACTGATGATGATGATGATGATGATG 815
Db 810 TCACATGCTTGTGTTTTCAGTAGTCTGACATGATGAGCTT---ATGAGCTCTATGTCAC 866

```

[illegible]

Db 452 TTCGCTTTAGGGGTTGCCATACCTGAATGTCAGAAAGCTATTACCAGGAGCGAAGCCCG 511
 QY 365 TTCGCTTCCGGGGTTGCTATACCTGAATGCCAAAGGATTACCTGCAGCAAGACCTG 424
 Db 512 GTGGAGAACCATTCCTGAGGGTCTGCTGGCTCTCTTTAACTGNAAGGTACCAACTA 571
 QY 425 GGGGTGAGCCCTTGCTGAAGGCTTCTCTGGCTCTTTTAAAGGAAAGGTGCCATCAA 484
 Db 572 AAGAGCAAGTAGATGCA-TTCTCGCGCAGAGTTGGAAAGTGGTGTGC-TGTACCAAGCA 629
 QY 485 AAGACCAAGTGAATCAATTCCTCAGGAATTCGAGAGTCGGGCATCATATCCTCATCA 544
 Db 630 TG-TATACAAACCATCATCATCGCTTACTGTGTACAGCTCATCCAAATGACATTTGCAA 688
 QY 545 TCATGTATACAACTATTGATGCTTACAGCTTACAGCTACAGCTCATCCAAATGACCAATTCGCTA 604
 Db 689 CTGGTGTATGCTGCTCAGCTTCAAGTGAATTCAGAGGATACGAGAAAGGATTC 748
 QY 605 CTGGAGTATGCTCTTACAGCTTCAAGTGAATTCAGAGGATACGAGAAAGGATTC 664
 Db 749 ACAAACCAAGTATTGGAGGCAACATATGAGGACTCTATTACTTTAAATTCCTCAATTAC 808
 QY 665 ACAATCAAGATTATGGGAACCAACATATGAGGATTCATGAATTCATTCCTCAAGTTC 724
 Db 809 CAGTTGTAGCAGCTTATATCTATCGAGGATGTACAGAAATGGACAAAGTATATCCACGG 868
 QY 725 CACTTGTGCTGCTTATGTTATCGCAGGATGTACAAAGATGTGACACTATACCTTAAGG 784
 Db 869 ATGATTCCTAGATTATGCTGCAACCTTTGCCACATGCTTGTGTAGATGATGCCAGCA 928
 QY 785 ATGAATCCTGATATGCTGCAAAATTTGCTACATGCTTGTTCAGTAGCTCTGAAA 844
 Db 929 TCGAAGAGCTT---ATGAGGCTTTACGTTACTATCCATCTACTACCAATGAGGTGGAAACG 985
 QY 845 TGCATGAACCTCTTATGAGGCTCTATGTAAACATACACAGTATCATGAAGGTGGTAAG 904
 Db 986 TTAGTGCTCATCTAGTCTAGTTCAGAGTGCCTTTTACAGACCATATCTTCTTTTG 1045
 QY 905 TCAGTGCTCAGACCGGTCACCTGTTGCTAGTGTGTTGCTGATCTTACCTCTCCTTTG 964
 Db 1046 CAGCTGCATTAAATGGTTTAGCAGGGCTCTTATGCTGTTGGCAATTCAGGAAGTTTGT 1105
 QY 965 CTGCTGCTTGAATGGTTTAGCGGACCACTTCATGCTTTAGCCATCAGGAAGTTTGC 1024
 Db 1106 TATGATCAAGTCTGTGTTTCCGAATGCGGAGAAATGTAACTAAGGAACAATGAAAG 1165
 QY 1025 TATGATAAATCTGTTAGAGAAATGTGGGAGAACATTTCCAAAGACGATTTGAAAG 1084
 Db 1166 ATTATATCTGGAACACATTAAACAGCGGAGGTTGTTCTCGATATGGACATGTTGTC 1225
 QY 1085 ACTATGTTTGGAAACATTAACAGTGGCAGAGTTGCTGCTGTTTGGACATGGAGTTC 1144
 Db 1226 TCGTAAACACATCCAGATACATATGTACAGAGAGTGTGCTTTAAAGCATTTACCTG 1285
 QY 1145 TCGAAAGACTGTACCAAGATATACATGCCAGAGAGATTCGCTATCAAGCATTTGCCG 1204
 Db 1286 ATGATCCACTCTTCAATGTTTCAACCTCTTTGAAGTGGTGGCTCCCTATCTTACAG 1345
 QY 1205 AAGATCCACTGTTCATGTTTCAAACTCTACGAAGTTT-CCTCCTGTTCTTACAG 1263
 Db 1346 AACTTGGCAAGTAAAGACCCCTGGCCAAATGTTGATGCCATAGTGGGGTACTGCTGA 1405
 QY 1264 AACTTGGCAAGTTAA-AACCT-TGGCCAAATGTTGATGCCACAGAGTGGTGTGTTGA 1321
 Db 1406 ATCATATGGGCTAACAGAACCCAGATATATATCTGTTCTTCTTGGAGTATCAGAGGCCA 1465
 QY 1322 ACTATTATGGTTAACTCAAGCAAGATATATACGGTCTCTTGGCGTATCAAGAGCTC 1381
 Db 1466 TTGGTATTGTCTCAGCTGTATGGGACAGAGCTCTTGTGATACCACTTGAGAGGCCAA 1525
 QY 1382 TTGGCATTTGCTCTCAGCTAATTTGGGACGAGCTCTTGTGATTCGCGCTAGAGAGGCCAA 1441
 Db 1526 AAAGTGTACAATGAATGCTTGAAGGCTTGAAGGACCTGCAAGAAATCTTCTTAAGTTTGA 1583

QY 1442 AGAGTGTCAAAATGGAGTGGCTTGAGAACCAAGTGCAGAAAGCATGAA-TGTTTGAAA 1499
 RESULT 6 CMU19481 1733 bp mRNA PLN 04-DEC-1995
 LOCUS Citrus maxima citrate synthase (cit) mRNA, nuc.ear gene encoding
 DEFINITION mitochondrial protein, complete cds.
 ACCESSION U19481
 NID 9624675
 VERSION U19481.1 GI:624675
 KEYWORDS pummelo.
 SOURCE Citrus maxima
 ORGANISM Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
 Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
 Magnoliopsida; Rutaceae; Sapindales; Rutaceae; Citrus.
 REFERENCE 1 (bases 1 to 1733)
 AUTHORS Canel,C., Bailey-Serres,J. and Roose,M.
 TITLE Physiological and molecular genetic studies of acid accumulation in
 JOURNAL citrus fruits
 Thesus (1994) Botany & Plant Sciences, University of California at
 Riverside 2 (bases 1 to 1733)
 REFERENCE Canel,C.
 AUTHORS Direct Submission
 TITLE Submitted (04-JAN-1995) Camilo Canel, University of California at
 JOURNAL Riverside, Botany & Plant Sciences, Riverside, CA 92521-0124, USA
 FEATURES
 source
 1..1733
 /organism="Citrus maxima"
 /strain="Siamese sweet (pummelo 2240)"
 /note="There is a single copy of the cit gene in the
 citrus genome. A single transcript with an estimated
 length of 1,850 nucleotides can be detected in juice
 cells."
 /db_xref="taxon:37334"
 /sex="hermaphrodite"
 /cell_type="juice cell"
 /tissue_type="juice tissue"
 /dev_stage="immature"
 5'UTR 1..56
 transit_peptide 57..111
 /gene="cit"
 /function="targets the protein to the mitochondrial
 matrix"
 gene 57..1472
 /gene="cit"
 CDS 57..1472
 /gene="cit"
 /codon_start=1
 /function="synthesis of citrate from oxaloacetate and
 acetylCoA"
 /product="citrate synthase precursor"
 /protein_id="AA82743.1"
 /db_xref="PID:9624676"
 /db_xref="GI:624676"
 /translation="MASLRSATLSRLSPAGQQSN:SNVFWLQMQSSADLDLHSQL
 KEMIPQERLKKVSDGLKAQLGNITIDVIGMGMTGLLWETSLDPDEGIPFG
 LSIPQCKLLPAKPDGPELPEGLWLLLTGKVPKSEOVGDLGSLKELDRATVPDYVK
 AIDALPVSAPMTQFASGVMAQVQSEFQEAKEGKHKSWEPTSESLNLAAPV
 VAAYVYRIYKDGKIIIPKDDSLDYGNSFSLMGFDPPKMLRLYVTHSDHEGNNV
 SATHGLVASALSDPYLSFLAALNGLAGPLHGLANGVLLWIKSVYDDECGNVTTEQL
 KDYVWKLNSGKWVFGHGVLRKTPRYTCOREFALKHLPDPLFOLVLSKLYEVP
 IITKLCKVKNPMPNVDHSGVLLNHFLGAELARYTYTLFGVSRSLGICQLIWDRLGL
 PLERPKSTLDWIENCKKAA"
 111..1469
 /gene="cit"
 /function="synthesis of citrate from oxaloacetate and
 acetylCoA"
 /product="citrate synthase"
 1473..1733
 1710..1716
 mat_peptide
 3'UTR
 polyA_signal

polyA_site 1733
BASE COUNT 451 a 324 c 446 g 512 t
ORIGIN

Query Match 37.0%; Score 700; DB 27; Length 1733;
Best Local Similarity 77.5%; Pred. No. 0.00e+00;
Matches 1105; Conservative 0; Mismatches 303; Indels 17; Gaps 12;

Db 52 GAACCATGGCGTCCCTCAGAGGGCCACTGGCGCTCTCTCGCCTCGGCTCTCGCGCTGGGC 111
Qy 68 GACCAATGGTGTCTACCTAGAGGTTTCGTTGCTCAAGCTCGGCTCTCGAGCGGTCC 127
Db 112 AACAGTCTAATCTCAGTAACTCAGTGGCTCAAGTCAAGTCAAGTCTCTGATCTTG 171
Qy 128 AACAGTCAATGTAGCAATCTGTGCGTGGCTCAAGTCAAGTCTCTGCGTCTTG 187
Db 172 ACCTTATCTCAGCTC -A-AGGAATGATTCCAGAACACAGAGCGCTCGTAAGAAG 228
Qy 188 ATCTGCGTTCGAGCTGGTACAAGAAATGATTCCTGAACACAGGATCGCTGAAAGA 247
Db 229 TGAAGTCAGACCTTGAAGAGCCAGCTGGGAATATTACCATTGACGTGTAATGGCG 288
Qy 248 TCAAGTCAGATAT -GAAAGGTTCAATGGGAAC-ATCACGTTGATGTTCTTGGTG 304
Db 289 GGATGAGAGAAATGACTGGCTTACTTTGGGAAACCTCATTAC-TTGACCCCTGATGAGGA 347
Qy 305 GAATGAGAGAAATGACAGGATTAATCTGG -AAACCTCATTACCTTGACCCCTGATGAGGA 363
Db 348 ATTGCGCTTCGGGGTCTGTCTATCTCGAGTGGCAGAGTGTGTTACCTGCGCGCTAAGCCA 407
Qy 364 ATTGCGCTTCGGGGTGTGCTATACCTGAATGCCAAAGGATTATACCTGCGACGAAGCCT 423
Db 408 GATGGAGAACCTTTGCTGAGGGTCTCTTGGCTCTTTTAACTGGAAGGTAACCGAGC 467
Qy 424 GGGGTGAGCCCTTGCTGAGAGTCTTCTGCTCTCTTTTAAAGAGAAAGTGCCATCA 483
Db 468 AAAGACAGCTCGATGCG -TTGTCCAAAGAAATGAGAGATCGTGCCAC-TGTGCCAGATT 525
Qy 484 AAAGAGCAAGTGAATCAATTTGCTCAGCAATTTGACAGATGGCAGCTCATATCCCTGATC 543
Db 526 ATG-TGTACAAGGCCATTGATGCTCTACCTGTATGGCAGATCCGATGACTCAGTTTCT 584
Qy 544 ATCATGATACAACTATTGATGCTTACCAGTCAGAGCTATCCAAATGACCAAGTTGCT 603
Db 585 AGTGGGTCATGCTCTGAGGTGCAAGTGAATTTGAGGAGCTTATGAGAGGGAAAT 644
Qy 604 ACTGGAGTCATGCTCTTCCAGGTTCAAGTGAATTTCAAAAGGCATACGAGAGGGATT 663
Db 645 CATAAATCAAGTCATGGGAACCAACATCTGAGGATTCCTTAATTTGATGCTGCTG 704
Qy 664 CACAAATCAAGTATTGGGAACCAACATATGAGGATTCATGAATCTGATGCTCAAGT 723
Db 705 CCGGTAGTACGATGATGTTTATCAAGGATATACAGAGTGAAGAAATATTCCCAAG 764
Qy 724 CCAGTTGCTGCTTATGTTTATCCAGAGTGTACAGAAATGGTGACATATACCTTAAG 783
Db 765 GATGATCTCAGTATGATGTTGGGAATCTTCCACATGTTGGGATTCGATGCTTAAG 824
Qy 784 GATGAATCCCTGATGATGTTGCAAAATTTGCTCAGATGCTTGGTTTCAGTAGCTCTGAA 843
Db 825 ATGCTAGAGCTT---ATGAGGCTTATGTTTACCATTCAGTGACCAAGAGGTGGAAT 881
Qy 844 ATGCTAGAGCTTCTTATGAGGCTTATGTTAACAATACAGTATGATCATGAGGTGGAAT 903
Db 882 GTCAGTGTCTACAGCGGCACTAGTTGCAAGTGGCACTTTCAGATCTTATCTTCTTTC 941
Qy 904 GTCAGTGTCTACAGCGGTCATGTTGTTGATGCTTGTCTGATGCTTACCTCTCTT 963
Db 942 TTGGCGCATTAATGGTTTACCTGGACCACTCCATGTTGGCTTAATCAGAGAGTTTG 1001
Qy 964 GCTGCTGCTTTGAATGATTTAGCGGACCACTTCATGTTTGGCTTATAGCAATCAGAGTTTG 1023
Db 1002 CTTTGGATCAAACTCTGTTGTTGATGATGTTGTTGAGAGCTGAACCAAGGAGCTGAA 1061

Qy 1024 CTATGATATAATCTGTTGTAGAGAAATGTGGGAGAAACATTTCCAAGAGCACTTGA 1083
Db 1062 GATTATGCTGAAACAACTTGAATAGTGGCAAGTGTGCTCTGGAATTTGGTATGATGTT 1121
Qy 1084 GACTATGTTTGGAAACAACTTGAACAGTGGCAAGTGTGCTCTGGAATTTGGACATGG 1143
Db 1122 TTGCGTAAACATGCTCAGAGATACACTTGTGAGAGGAGTTTGGCTTGAAGCATTTG 1181
Qy 1144 CTGCGAAAGCACTGTACCAAGATATACATGCCAGAGAGATTCGCTATCAAGCATTT 1203
Db 1182 GATGATCCATTTTCCAGCTGCTTTTAAAGCTGTATGAAGTGTGCTGCTGCTTAAACA 1241
Qy 1204 GAAGATCCACTGTTTCAACTGCTTCAAACTCTACGAAGTTCCTCTCTCTTCTTCA 1262
Db 1242 AAAGTAGGCAAGGTTTAAAGCCCTGCGCAAAATGTTGATGCTCAGAGTGGGCTGCTG 1301
Qy 1263 GAAGTGGCAAGTGTAAAA--CCTTGGCAAAATGTTGATGCCACAGTGTGTTGTTG 1340
Db 1302 AACCATTTTGGTGTGCTGAGCAAGATATTATCTGCTCTTTTGGAGTATCAAGGAT 1361
Qy 1321 AACTATATGTTTAACTGAAGCAAGATATTATGCTGCTCTTTGGCTATCAAGAGCT 1380
Db 1362 CTGGGATCTGCTCTCAGTGTATGATGGACGAGCTCTTGGCTGCTGCTGCTGAGAGCA 1421
Qy 1381 CTGGGATTTGCTCTCAGCTAATTTGGACCGAGCTCTTGGATTTGCTGCTGAGAGGCA 1440
Db 1422 AAAGTGTAACTTTAGATTGATTCGAGAAAATTTGCAAGAAAGCA 1466
Qy 1441 AAAGTGTCAAAATGAGTGGCTTGGAGAACCAAGTGTCAAGAAAGCA 1485

7 A46546 1551 bp DNA PAT 07-MAR-1997
LOCUS Sequence 2 from Patent WO9524487.
DEFINITION A46546
ACCESSION 92300709
NID 92300709
VERSION A46546.1 GI:2300709
KEYWORDS
SOURCE beet.
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Caryophyllales; Caryophyllales; Chenopodiaceae; Beta.
1 (bases 1 to 1551)
Muehler-Roeber, B., Landschuetze, V. and La, C. U.
PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN
PLANTS
Patent: WO 9524487-A 2 14-SEP-1995;
HOECHST SCHERING AGREVO GMBH (DE)
Other publication DE 4438821 960425
Other publication DE 4435366 960328
Other publication AU 2067995 950925
Other publication DE 4408629 950914.
FEATURES
source
1..1551
/organism="Beta vulgaris"
/strain="ZUCHTLINIE '5S 0026"
/db_xref="taxon:3555"
/tissue_type="LEAF"
/clone="PSBCS"
/EC_number="4.1.3.7"
/note="Protein sequence is in conflict with the conceptual
translation."
/codon_start=1
/product="GIRATE SYNTHASE"
/protein_id="CAA02909.1"
/db_xref="PID:e306423"
/db_xref="PID:g2300710"
/db_xref="GI:2300710"
/translation="SSNLDRLSELQELIPEQEPKIKKKEFSPLSNINIMVLGG
MRGWTGLWETSLLDPDEGIRFRGFSIPEQCKLLPAASAGAEPLPELLWLLITKVP

SKQVDALSADLRKRASIPDHVYKTIIDALPITAHPTMTQCTGVMAOTRSEFOKAYEK
 GHKSEWPEPTEDCLSLIAQVPPVAAVYVYRMKNGQVPLDDSLDYGNGFAMHLE
 DSPMILEMLRLXVTIHSDEHGNVSAHTGLVGLSPDYLSPFAAALNGLAGPLHGLA
 NORVLLWIKSVNDEGENISTEQLKDYVWKTILNSGVVPGFGLGVLPKTDPPPTCQRE
 FALKHLPDFFQFVSKLVEVPPILLELGVKPNPNDVDAHSGVLLNHLGLTEARYI
 TVLFGVRSGLGICSLINDRALGLPLERPKPVTMEWLEKFKRRA*

BASE COUNT 455 a 268 c 370 g 458 t
 ORIGIN

Query Match 35.18; Score 664; DB 25; Length 1551;
 Best Local Similarity 78.58; Pred. No. 0.00e+00;
 Matches 997; Conservative 0; Mismatches 261; Indels 12; Gaps 10;

Db 29 TACAAGAACTGATTCCTGAACAACAGAGAGCTGAAGAAGATAAAGAAAGAAATTGGAA 88
 Qy 206 TACAAGAATTGATTCCTGAACAACAGAGATCGCTGAAAGAAATCAAGTCAGA-TATGAAA 264

Db 89 GTTTCACCTGGGGAATATCAATGTTGACATGTTATGGCGGAATGAGAGAAAGTACTG 148
 Qy 265 GGTTCAAT-TGGG-AACATCACAGTTGATATGTTCTTGGTGAATGAGAGGAATGACAG 322

Db 149 GTTTACTTTGGGAGACTTCCTTACTCGACCCAGCAAGAGGATATCCGGTTCAGGGGTTTTT 208
 Qy 323 GATTACTGTGGAACCTCATACCTTGACCTGATGAGGAATTCCTTCCGGGGGTGT 382

Db 209 CTATACCTGAATGCCAGAAACTTTTACCCCGCTGCAAGTGTGTCGAGAGCCATTCGCTG 268
 Qy 383 CTATACCTGAATGCCAAAGGATATTACCTGCAGCAAGGCTGGGGGTGAGCCCTTGCCTG 442

Db 269 AAGCTCTTCTTGGCTTTTAAACCGGAAGGTTCTTAGCAAGAGCAAGTAGATGCTC 328
 Qy 443 AAGCTCTTCTTGGCTTTTAAACCGGAAGGTTCTTAGCAAGAGCAAGTAGATGCTC 502

Db 329 TATCAGCAG-ATTTCAGAAAACGTGCTTC-TATCCCGAGACCATG-TGTACAAACAATG 385
 Qy 503 TTGCTCAGGAATTGCGAGTTCGGGATCATATCCCTGATCATCATGATATACAATATG 562

Db 386 ATGCTCTACCTATTACGGCTATCCCAATGACTCAGTTTTTGGCTGCTGTTATGGCTTAC 445
 Qy 563 ATGCTTACCAGTACAGCTCATCCCAATGACCCAGTTTGTACTGGAGTCATGGCTCTC 622

Db 446 AGACTCGAAGGAAATTCAGAAAGCATATCAGAAAGGATCCATAGTCAAAAGTTTGGG 505
 Qy 633 AGTTCAAGTGAATTTCAAAGGCATACAGAAAGGATTCACAAATCAAGATATGG 682

Db 506 AGCCAAATATGAGSACTGCTTATGTTGATGCTCAAGTTCCTGTTGTCAGCTTATG 565
 Qy 683 AACCAACATATGAGGATTCCTCAATCTGATGCTCAAGTTCCACTTTGTTGCTGCTTATG 742

Db 566 TTTATCGGAGATGATAGAATGGACAAGTAAATACCGTGGATGATGACTCCCTGATTATG 625
 Qy 743 TTTATCGGAGATGATAGAATGGTGGACACTATACCTAAGGATGAATCCCTGGATTATG 802

Db 626 GTGAAATTTCCACATATGTTGGATTTTCATAGCCCTCAGATGCTTGGAGCTG---ATGC 682
 Qy 803 GTGCAATTTTGTCCATGCTGTTGTTTCAAGTCTCTGAAATGCAATGATCTTATGA 862

Db 683 GCCTTTATGTCACAAATTCACAGTATCATAGGAGTGAATGTTAGTGCACACTGGCC 742
 Qy 863 GGCCTATGTAAACATACAGTATCATGAAGTGGTATGTCAGTGTCTCACACCGGTC 922

Db 743 ATTTGGTGGTATGTCACCTTTCAGATCCTTATTTGTCATTTGCGAGCAGCAATTAATGTT 802
 Qy 923 ACTTGGTGTGATGCTGTTGCTGATCCTTACCTCTCCTTGTGCTGCTTTCATGTT 982

Db 803 TGGCTGGGCCACTCCATGGATTAGCCCAACAGGAGTCTGCTGCTGATTAAATCAGTTG 862
 Qy 983 TAGCCGGACCACTTCATGGTTTAGCCCAATCAGGAAGTTTGTATGATGAATATCTGTTG 1042

Db 863 TTGATGAATGTGGAGAACATCTCGACAGAGCAGTTGAAAGATATATGTTTGGAGACAC 922
 Qy 1043 TAGAAGATGTGGGGAACATTTCCCAAGAGCAGTTGAAAGACTATGTTTGGAAACAT 1102

Db 923 TAAACAGTGGCAAGGTTGTACCTGGATTGGTCTAGGAGTATTGGCAAGACAGATCCAA 982
 Qy 1103 TGACAGTGGCAAGGTTGTCCCTGGTTTGGACATGGAGTCTGCGAAGACTGTACCAA 1162

Db 983 GATACACATGGCCAAAGAGAAATTTGGTGTGAACACTTGCCTGATGACCCATTTTTCAT 1042
 Qy 1163 GATATACATGCCAGAGAGATTCGCTATGAGCAATTTGCTGGAAGATCCACTGTTTCAAC 1222

Db 1043 TGTGTGTCAAAGTGTATGAAGTGGTGGCTCCCTTATTCTATTAGAGCTTGGAAAGTTAAGA 1102
 Qy 1223 TGGTTTCAAAACTCTACGAAGTTT-CTTCTGTTCTTTACAGAAGTTTGGCAAGTTAA-A 1280

Db 1103 ATCCATGGCTTATGTTGATGCTCATAGTGGAGTTTTCGTGAACCACTATGTTTACAG 1162
 Qy 1281 A-CCTTGGCCAAATGTTGATGCCACAGTGTGTGTTGTTGAACATATTATGTTTAACTG 1339

Db 1163 AAGCAAGACTACTACGGTTTCTTGGGCTATCAAGAGTCTTGGAAATATGCTCACAGC 1222
 Qy 1340 AAGCAAGATATTATACGGTCTTTCGCTGATCAAGAGCTCTTGGCATTTGCTCTCAGC 1399

Db 1223 TTATATGGGACCGAGCTCTTGGCTTGGCTTATCAAGAGCTCTTGGCATTTGCTCTCAGC 1282
 Qy 1400 TAATTTGGGACCGAGCTCTTGGATTGCCCTAGAGAGCCAAAGAGTGTACATATGGAGT 1459

Db 1283 GGCTTCAAAA 1292
 Qy 1460 GCCTTGAGAA 1469

RESULT 8

LOCUS BVCITSYNH 1551 bp mRNA PLN 07-OCT-1996
 DEFINITION B.vulgaris mRNA for citrate synthase.
 ACCESSION X84228

NID g1556379

VERSION X84228.1 GI:1556379

KEYWORDS citl gene; citrate synthase.

SOURCE beet.

ORGANISM

Beta vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.

REFERENCE

1 (bases 1 to 1551)
 La Cognata,U., Landschuetz,V., Willmitzer,L. and Mueller-Roeber,B.
 Plant Cell Physiol. In press

AUTHORS

2 (bases 1 to 1551)
 La Cognata,U.
 Direct Submission
 Submitted (23-JAN-1995) U. La Cognata, Institute fuer
 Genbiologische Forschung, Berlin GmbH, Ihnestrasse 63, 14195
 Berlin, FRG

FEATURES

Location/Qualifiers
 1..1551
 /organism="Beta vulgaris"
 /db_xref="taxon:3555"
 1..1314
 /gene="citl"
 <1..1314
 /gene="citl"
 /EC_number="4.1.3.7"
 /codon_start=1
 /product="citrate (sl)-synthase"
 /protein_id="CAAS9010.1"
 /db_xref="PID:e137433"
 /db_xref="PID:g1556380"
 /db_xref="GI:1556380"
 /db_xref="SPTREMBL:O96544"
 /translation="SSNLDLRSEQLLEPEOERLKKIKKEFSFOLGNINVMVLGG
 MRMTGLLWETSLLDPBEGIRGRFSIPECKLLPAASAGAEPLPEGLLLLLLTGKVP
 SKEQVDALSADLRKRASIPDHVYKTIIDALPITAHPTMTQCTGVMAOTRSEFOKAYEK
 GHKSEWPEPTEDCLSLIAQVPPVAAVYVYRMKNGQVPLDDSLDYGNGFAMHLE
 DSPMILEMLRLXVTIHSDEHGNVSAHTGLVGLSPDYLSPFAAALNGLAGPLHGLA
 NORVLLWIKSVNDEGENISTEQLKDYVWKTILNSGVVPGFGLGVLPKTDPPPTCQRE
 FALKHLPDFFQFVSKLVEVPPILLELGVKPNPNDVDAHSGVLLNHLGLTEARYI
 TVLFGVRSGLGICSLINDRALGLPLERPKPVTMEWLEKFKRRA"


```

BASE COUNT 455 a 268 c 370 g 438 t
ORIGIN
    1 TVLPGVSRSLGICGSLIWDRAGLPLERPKSVTMWLEKECKBPA*
    2
    3
    4
    5
    6
    7
    8
    9
    10
    11
    12
    13
    14
    15
    16
    17
    18
    19
    20
    21
    22
    23
    24
    25
    26
    27
    28
    29
    30
    31
    32
    33
    34
    35
    36
    37
    38
    39
    40
    41
    42
    43
    44
    45
    46
    47
    48
    49
    50
    51
    52
    53
    54
    55
    56
    57
    58
    59
    60
    61
    62
    63
    64
    65
    66
    67
    68
    69
    70
    71
    72
    73
    74
    75
    76
    77
    78
    79
    80
    81
    82
    83
    84
    85
    86
    87
    88
    89
    90
    91
    92
    93
    94
    95
    96
    97
    98
    99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
    701
    702
    703
    704
    705
    706
    707
    708
    709
    710
    711
    712
    713
    714
    715
    716
    717
    718
    719
    720
    721
    722
    723
    724
    725
    726
    727
    728
    729
    730
    731
    732
    733
    734
    735
    736
    737
    738
    739
    740
    741
    742
    743
    744
    745
    746
    747
    748
    749
    750
    751
    752
    753
    754
    755
    756
    757
    758
    759
    760
    761
    762
    763
    764
    765
    766
    767
    768
    769
    770
    771
    772
    773
    774
    775
    776
    777
    778
    779
    780
    781
    782
    783
    784
    785
    786
    787
    788
    789
    790
    791
    792
    793
    794
    795
    796
    797
    798
    799
    800
    801
    802
    803
    804
    805
    806
    807
    808
    809
    810
    811
    812
    813
    814
    815
    816
    817
    818
    819
    820
    821
    822
    823
    824
    825
    826
    827
    828
    829
    830
    831
    832
    833
    834
    835
    836
    837
    838
    839
    840
    841
    842
    843
    844
    845
    846
    847
    848
    849
    850
    851
    852
    853
    854
    855
    856
    857
    858
    859
    860
    861
    862
    863
    864
    865
    866
    867
    868
    869
    870
    871
    872
    873
    874
    875
    876
    877
    878
    879
    880
    881
    882
    883
    884
    885
    886
    887
    888
    889
    890
    891
    892
    893
    894
    895
    896
    897
    898
    899
    900
    901
    902
    903
    904
    905
    906
    907
    908
    909
    910
    911
    912
    913
    914
    915
    916
    917
    918
    919
    920
    921
    922
    923
    924
    925
    926
    927
    928
    929
    930
    931
    932
    933
    934
    935
    936
    937
    938
    939
    940
    941
    942
    943
    944
    945
    946
    947
    948
    949
    950
    951
    952
    953
    954
    955
    956
    957
    958
    959
    960
    961
    962
    963
    964
    965
    966
    967
    968
    969
    970
    971
    972
    973
    974
    975
    976
    977
    978
    979
    980
    981
    982
    983
    984
    985
    986
    987
    988
    989
    990
    991
    992
    993
    994
    995
    996
    997
    998
    999
    1000
    1001
    1002
    1003
    1004
    1005
    1006
    1007
    1008
    1009
    1010
    1011
    1012
    1013
    1014
    1015
    1016
    1017
    1018
    1019
    1020
    1021
    1022
    1023
    1024
    1025
    1026
    1027
    1028
    1029
    1030
    1031
    1032
    1033
    1034
    1035
    1036
    1037
    1038
    1039
    1040
    1041
    1042
    1043
    1044
    1045
    1046
    1047
    1048
    1049
    1050
    1051
    1052
    1053
    1054
    1055
    1056
    1057
    1058
    1059
    1060
    1061
    1062
    1063
    1064
    1065
    1066
    1067
    1068
    1069
    1070
    1071
    1072
    1073
    1074
    1075
    1076
    1077
    1078
    1079
    1080
    1081
    1082
    1083
    1084
    1085
    1086
    1087
    1088
    1089
    1090
    1091
    1092
    1093
    1094
    1095
    1096
    1097
    1098
    1099
    1100
    1101
    1102
    1103
    1104
    1105
    1106
    1107
    1108
    1109
    1110
    1111
    1112
    1113
    1114
    1115
    1116
    1117
    1118
    1119
    1120
    1121
    1122
    1123
    1124
    1125
    1126
    1127
    1128
    1129
    1130
    1131
    1132
    1133
    1134
    1135
    1136
    1137
    1138
    1139
    1140
    1141
    1142
    1143
    1144
    1145
    1146
    1147
    1148
    1149
    1150
    1151
    1152
    1153
    1154
    1155
    1156
    1157
    1158
    1159
    1160
    1161
    1162
    1163
    1164
    1165
    1166
    1167
    1168
    1169
    1170
    1171
    1172
    1173
    1174
    1175
    1176
    1177
    1178
    1179
    1180
    1181
    1182
    1183
    1184
    1185
    1186
    1187
    1188
    1189
    1190
    1191
    1192
    1193
    1194
    1195
    1196
    1197
    1198
    1199
    1200
    1201
    1202
    1203
    1204
    1205
    1206
    1207
    1208
    1209
    1210
    1211
    1212
    1213
    1214
    1215
    1216
    1217
    1218
    1219
    1220
    1221
    1222
    1223
    1224
    1225
    1226
    1227
    1228
    1229
    1230
    1231
    1232
    1233
    1234
    1235
    1236
    1237
    1238
    1239
    1240
    1241
    1242
    1243
    1244
    1245
    1246
    1247
    1248
    1249
    1250
    1251
    1252
    1253
    1254
    1255
    1256
    1257
    1258
    1259
    1260
    1261
    1262
    1263
    1264
    1265
    1266
    1267
    1268
    1269
    1270
    1271
    1272
    1273
    1274
    1275
    1276
    1277
    1278
    1279
    1280
    1281
    1282
    1283
    1284
    1285
    1286
    1287
    1288
    1289
    1290
    1291
    1292
    1293
    1294
    1295
    1296
    1297
    1298
    1299
    1300
    1301
    1302
    1303
    1304
    1305
    1306
    1307
    1308
    1309
    1310
    1311
    1312
    1313
    1314
    1315
    1316
    1317
    1318
    1319
    1320
    1321
    1322
    1323
    1324
    1325
    1326
    1327
    1328
    1329
    1330
    1331
    1332
    1333
    1334
    1335
    1336
    1337
    1338
    1339
    1340
    1341
    1342
    1343
    1344
    1345
    1346
    1347
    1348
    1349
    1350
    1351
    1352
    1353
    1354
    1355
    1356
    1357
    1358
    1359
    1360
    1361
    1362
    1363
    1364
    1365
    1366
    1367
    1368
    1369
    1370
    1371
    1372
    1373
    1374
    1375
    1376
    1377
    1378
    1379
    1380
    1381
    1382
    1383
    1384
    1385
    1386
    1387
    1388
    1389
    1390
    1391
    1392
    1393
    1394
    1395
    1396
    1397
    1398
    1399
    1400
    1401
    1402
    1403
    1404
    1405
    1406
    1407
    1408
    1409
    1410
    1411
    1412
    1413
    1414
    1415
    1416
    1417
    1418
    1419
    1420
    1421
    1422
    1423
    1424
    1425
    1426
    1427
    1428
    1429
    1430
    1431
    1432
    1433
    1434
    1435
    1436
    1437
    1438
    1439
    1440
    1441
    1442
    1443
    1444
    1445
    1446
    1447
    1448
    1449
    1450
    1451
    1452
    1453
    1454
    1455
    1456
    1457
    1458
    1459
    1460
    1461
    1462
    1463
    1464
    1465
    1466
    1467
    1468
    1469
    1470
    1471
    1472
    1473
    1474
    1475
    1476
    1477
    1478
    1479
    1480
    1481
    1482
    1483
    1484
    1485
    1486
    1487
    1488
    1489
    1490
    1491
    1492
    1493
    1494
    1495
    1496
    1497
    1498
    1499
    1500
    1501
    1502
    1503
    1504
    1505
    1506
    1507
    1508
    1509
    1510
    1511
    1512
    1513
    1514
    1515
    1516
    1517
    1518
    1519
    1520
    1521
    1522
    1523
    1524
    1525
    1526
    1527
    1528
    1529
    1530
    1531
    1532
    1533
    1534
    1535
    1536
    1537
    1538
    1539
    1540
    1541
    1542
    1543
    1544
    1545
    1546
    1547
    1548
    1549
    1550
    1551
    1552
    1553
    1554
    1555
    1556
    1557
    1558
    1559
    1560
    1561
    1562
    1563
    1564
    1565
    1566
    1567
    1568
    1569
    1570
    1571
    1572
    1573
    1574
    1575
    1576
    1577
    1578
    1579
    1580
    1581
    1582
    1583
    1584
    1585
    1586
    1587
    1588
    1589
    1590
    1591
    1592
    1593
    1594
    1595
    1596
    1597
    1598
    1599
    1600
    1601
    1602
    1603
    1604
    1605
    1606
    1607
    1608
    1609
    1610
    1611
    1612
    1613
    1614
    1615
    1616
    1617
    1618
    1619
    1620
    1621
    1622
    1623
    1624
    1625
    1626
    1627
    1628
    1629
    1630
    1631
    1632
    1633
    1634
    1635
    1636
    1637
    1638
    1639
    1640
    1641
    1642
    1643
    1644
    1645
    1646
    1647
    1648
    1649
    1650
    1651
    1652
    1653
    1654
    1655
    1656
    1657
    1658
    1659
    1660
    1661
    1662
    1663
    1664
    1665
    1666
    1667
    1668
    1669
    1670
    1671
    1672
    1673
    1674
    1675
    1676
    1677
    1678
    1679
    1680
    1681
    1682
    1683
    1684
    1685
    1686
    1687
    1688
    1689
    1690
    1691
    1692
    1693
    1694
    1695
    1696
    1697
    1698
    1699
    1700
    1701
    1702
    1703
    1704
    1705
    1706
    1707
    1708
    1709
    1710
    1711
    1712
    1713
    1714
    1715
    1716
    1717
    1718
    1719
    1720
    1721
    1722
    1723
    1724
    1725
    1726
    1727
    1728
    1729
    1730
    1731
    1732
    1733
    1734
    1735
    1736
    1737
    1738
    1739
    1740
    1741
    1742
    1743
    1744
    1745
    1746
    1747
    1748
    1749
    1750
    1751
    1752
    1753
    1754
    1755
    1756
    1757
    1758
    1759
    1760
    1761
    1762
    1763
    1764
    1765
    1766
    1767
    1768
    1769
    1770
    1771
    1772
    1773
    1774
    1775
    1776
    1777
    1778
    1779
    1780
    1781
    1782
    1783
    1784
    1785
    1786
    1787
    1788
    1789
    1790
    1791
    1792
    1793
    1794
    1795
    1796
    1797
    1798
    1799
    1800
    1801
    1802
    1803
    1804
    1805
    1806
    1807
    1808
    1809
    1810
    1811
    1812
    1813
    1814
    1815
    1816
    1817
    1818
    1819
    1820
    1821
    1822
    1823
    1824
    1825
    1826
    1827
    1828
    1829
    1830
    1831
    1832
    1833
    1834
    1835
    1836
    1837
    1838
    1839
    1840
    1841
    1842
    1843
    1844
    1845
    1846
    1847
    1848
    1849
    1850
    1851
    1852
    1853
    1854
    1855
    1856
    1857
    1858
    1859
    1860
    1861
    1862
    1863
    1864
    1865
    1866
    1867
    1868
    1869
    1870
    1871
    1872
    1873
    1874
    1875
    1876
    1877
    1878
    1879
    1880
    1881
    1882
    1883
    1884
    1885
    1886
    1887
    1888
    1889
    1890
    1891
    1892
    1893
    1894
    1895
    1896
    1897
    1
```

Qy	72	AATGGTGTCTTACCGTAGCGTTCTGGTCTGTCTGTCATGAAGCTCCGGCTCTCGAG--C-C-GGTCCA	178
Db	117	ACAATCTTCACTCAGCAAAATTTCTGTCCAGATGGATTCCAGATGCAGAGCTCTACCGACCTGGA	176
Qy	129	ACAGTCAAAATGTTAGCAATTTCTGGCGTGGCTTCAAGTCCAAACCTCTTCCGGTCTTGA	188
Db	177	CTTGAAGTCCGACGTC--CAAGAGTTAAATTCGGGAACACAAGACACGCTCTGAAGAACT	233
Qy	189	TCCTGGCTCTGAGCTGGTACAAGAATTGATCTCTGAACAACAGATCGCTGAAAAAGAT	248
Db	234	GAAGTCAGAACATGGGAAGGTCCTCACTGGGAACATCACTGTTCATATGGTTATTTGGTGG	293
Qy	249	CAAGTCAGAT-ATGA-AAGGTTCAATTGGGA-CATCACAGTTGATGTTCTTGGTGG	305
Db	294	AATGAGAGGATGACTGGATTCGTTTGGGAACCTCATTGC-TTGACCCGGAAGAGTA-	351
Qy	306	AATGAGAGAAATGACAGGAATTACTGTGG-AAACCTCAATTAGCTTCACTGATGAGGAA	364
Db	352	TTCCGCTTTAGGGATGTGCACTTCCTGAGTGCCAGAAGCCCTATTACCTACTGACGCCAG	411
Qy	365	TTCCGCTCCGGGGTTCTCTA-TACTGNAITGCCA-AAAGG--TATTACCTGCAACAAG	420
Db	412	CTTGGAGGCTTGAACCAATTACCGGAGGCTTTTGTGGCTTCTCTTAACCTGGAAGGTACC	471
Qy	421	CTTGGGGGTGAGC-CTCTGGCTGAAGGCTTCTCTGGGCTCTCTTTAAACAGGAAAGTGGC	479
Db	472	TTAGCAAGAGCAAGTTGAAGCACT-GTCGAAGACTTTGGA-AC-CGTGTGC-TGTGTC	527
Qy	480	ATCA-AAAGAGCAAGTGAATTCATTTGCTCAGGAATTCAGAGTGGGCAATCATATCCC	538
Db	528	AGATTATG-TGTACAAATGCCATCGATGCTCTGCCCTCCACAGCTCATCCAACTGACTCAAT	586
Qy	539	TGATCATCATGTATACAACATATTGATGCGTTACCAGTCCACAGCTCATCCAAATGACCCAGT	598
Db	587	TTCCTAGCGGTGTTATGGCCCTCCAGGTCCAAGTGAAGTTCAAAAGCCATATGAGAATG	646
Qy	599	TTGCTACTGGAGTCATGGCTCTTCAGGTTCAAGTGAATTTCAAAGAGCATACAGAGAAG	658
Db	647	GAATTCATAAGTCAAAGTTCTGGGAGCCAAACATATGAGGATTTGCTCAACCTGATTTGCTC	706
Qy	659	GGATTCACAAATCAAAGTATTTGGGAACCAACATATGAGGATTTCCATGAATCTGATTCCTC	718
Db	707	CTGTTCTCTGTTGAGTCGATATGTTATTCGAAGAGATGTATAGAATGGTGATTCATTC	766
Qy	719	AAGTTCACACTGTGTGCTGTTATGTTTATTCGAGAGATGTACAAGATGTTGACACTATAC	778
Db	767	CCTCAGATAAATCTTTGGATATGTTGTGCAAAATTTTCCACATGTTGGGATTTTCAT-GAT	825
Qy	779	CTAAGGATGAAATCCCTGGATATGTTGCAAAATTTTGCTTCACATGCTGTGTTTCAGTACCT	838
Db	826	--GAAAGGTTGAAGAGCTCATGAGCT-TACATCACCATCCACAGTATGATGCATGAAGGT	882
Qy	839	CTGAAATGCAATGAACATCTTATGAGGCTCTATCTAACAAATACACAGTGAT-CATGAAGGT	897
Db	883	GGAAATGTTAGTCTCACACTGTGCACCTGGTTGGTGTAGTGACCTTTTCAGACCCCATCTG	942
Qy	898	GGTAATGTCAGTCTCACACCGGTCACTTGGTTGCTAGTGCTTTGTCGTGATTCCTTACCTC	957
Db	943	TCATTTCAGCTGCATTAATGGTTTATGCTGGGCCACTCCATGGTTTGGCTTAATCAGGAA	1002
Qy	958	TCCTTTGCTGCTGCTTGAATGGTTTAGCCGACCACTTCAATGGTTTAGCCAAATCAGGAA	1017
Db	1003	GTTTTGGTTTGGATCAAACTGCTAGAGGAATCTGGAGAGATATATCAAAAGACAG	1062
Qy	1018	GTTTGTGCTATGGATAAAATCTGTTGTAGAAGAAATGTGGGAGAACATTTTCCAAAGACAG	1077
Db	1063	TTGAAAGAATATGTTTGGAAAACATTAAACAGTGGCAAGGTTATTCGGGGATATGGTCAAC	1122
Qy	1078	TTGAAAGACATGTTTGGAAAACATTGAACAGTGGCAAGGTTTCCCTGGTTTGGACAT	1137
Db	1123	GGTGTCTTCGCCAATACTGATCCAGATATGATGCCAAAGAAATTTGCCCTTGAAGCAT	1182
Qy	1138	GGAGTTCTGGGAAAGACTGTAGCAAGATATACATGTCAGAGAGAGTTCGTATGAGCAT	1197

```

RESULT 10
LOCUS      PHCITSYNH      1378 bp      mRNA
DEFINITION Populus hybrid mRNA for citrate synthase.
ACCESSION  X84227
NID        GL556430
VERSION    X84227.1      GI:1556430
KEYWORDS   citl gene; citrate synthase.
SOURCE     Populus balsamifera subsp. trichocarpa X Populus deltoides.
ORGANISM   Populus balsamifera subsp. trichocarpa X Populus deltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Violales; Salicaceae; Populus.
REFERENCE  1 (bases 1 to 1378)
            La Cognata,U , Landschuetz,V., Willmitzer,L and Mueller-Roeber,B.
            JOURNAL Plant Cell Physiol. In press
REFERENCE  2 (bases 1 to 1378)
            La Cognata,U
            Direct Submission
            Submitted (23-JAN-1995) U. La Cognata, Institute fuer
            Genbiologische Forschung, Berlin GmbH, Ihnestrasse 63, 14195
            Berlin, FRG
FEATURES             Location/Qualifiers
     source           1..1378
                     /organism="Populus balsamifera subsp. trichocarpa X
                     Populus deltoides"
                     /note="hybrid"
     gene             /db_xref="taxon:3695"
                     16..1035
                     /gene="citl"
     CDS              16..1035
                     /gene="citl"
                     /EC_number="4.1.3.7"
                     /codon_start=1
                     /transl_except=(pos:16..18,aa:Met)
                     /product="citrate (si)-synthase"
                     /protein_id="CA459009.1"
                     /db_xref="PID:gi76838"
                     /db_xref="PID:gi1648926"
                     /db_xref="GI:1648926"
                     /db_xref="SPTREMBL:O24259"
                     /translation="MPTKEQVGALSKELRDRLVHDVYFKALDLPVTAHPMTQFATG
                     VMAQVQSEFKAYEKGIHKSXWPTEDSLIARPIVASIYIRYIKDKVPIPM
                     FDSLVIGNSHMLGDSPEMQLMRLVINSHSGEGNVSAHTGHVLASALSDPILS
                     NAAALNLAQLPHLGANVLLWIKSWEECENITTEQLKDYVMKTLNSGKVPVPGF
                     HGVLKRTYPRYTQREFALKHLPDDLPFLQSKLYEVYPPVLTQLGKVKPWPNDVAH
                     SGVLLNYGTLTEARYTYTVLFGVSRSIGCSQLIMDRALGLPLERPKSVTWELLENHCK
                     KAPAN"
BASE COUNT      396 a      252 c      313 g      417 t
ORIGIN
Query Match      29.8%; Score 564, DB 27; Length 1378.

```

LOCUS	AF047042	1401 bp	mRNA		PFI	05-JUL-1998
DEFINITION	Homo sapiens citrate synthase mRNA, complete cds.					
ACCESSION	AF047042					
NID	g3288814					
VERSION	AF047042.1	GI:3288814				
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 1401) Goldenthal,M.J. Cloning and sequence analysis of human citrate synthase cDNA Unpublished					
AUTHORS	2 (bases 1 to 1401) Goldenthal,M.J. Direct Submission Submitted (06-FEB-1998) Molecular Cardiology Institute, 75 Raritan Avenue, Highland Park, NJ 08904, USA					
JOURNAL	Location/Qualifiers					
FEATURES	1..1401 /organism="Homo sapiens" /db_xref="taxon:9606" /sex="male" /tissue_type="heart" /note="isolated from a 22-year old"					
CDS	1..1401 /codon_start=1 /product="citrate synthase" /protein_id="AAC25560.1" /db_xref="PID:g3288815" /db_xref="GI:3288815" /translation="MALLTAAARLLGTTKNASCLVLAARHASASTNLKDLADLPKE QAPLKTFRQOHGKTVGQITVDMYGGMPGMKGLVETSVLDPEGIRPFGPSIPFGQ KLHPKSGEELPEGLFWLLVTCIPTEQVSWLSKEWAKRAALPSHVVMIDNP NLHPMSQLSAATALNSENPAQYARGISFTKIYELIYEDSVDLAKLPVCAAKIYR NLYWEGSIGAIDNLMHSHNFTNMLGYTDHQTLEMLRXYLTIHSDHEGNGVSAHTSH LVGSLSDPYLSFAAMNGLAGPLHGLANGELVWLTOLEKVGKDYSDSKLDYIWN TLNSRVVPVGYHVALRKTDPRYTCOREFALKHLPNDPMEKLVQIYKIPNVLLQEQ KAKPNPNVDARHSGVLLQYYGEMTMYVTVLFGVSRALGVLAOLINSRALGFPIERP SMSTEGMLMKFVDSKSG"					
BASE COUNT	352 a	338 c	371 g	340 t		
ORIGIN						
Query Match	12.7%; Score 240; DB 31; Length 1401;					
Best Local Similarity	63.7%; Pred. No. 2,77e-187;					
Matches	746; Conservative u; Mismatches 410; Indels 1b; Gaps 13;					
Db	187	ATCACTGTGGACATGATGTATGGTGGCATGAGAGGCATGAGAGGATTTGGTGTATGAACA	246			
Qy	280	ATCACAGTTGATATGGTCTTGGTGAATGAGAGGAATGACAGGATTACTGTG-CAAAAC	338			
Db	247	TCAGTTC-TTGATCTGTATGAGGCGATCCGTTTCGAGGCGCTTTAGTATCCCTGAATGCCA	305			
Qy	339	TCATTACCTTGACCTGTATGAGGGAATTCGCTTCGCGGGGTGTCTATACCTGAATGCCA	398			
Db	306	GAACCTGTACCCAGGCTAAGGTGGGGAAGAACCCCTGCCTCAGGCGCTTATTTGGCT	365			
Qy	399	AAAGGTATTACCTGCAGCAAAAGCCCTGGGGGTGAGCCCTTGCTCTCTCTGCT	458			
Db	366	GTGTGTAAGTGGATGTATCCCAACAGAGAACAGGTATCTTGGCTC-TCAAAGAGTGGG	424			
Qy	459	TCCTTTAACAGGAAGGTGGCCATCAAAAGACGACAGTGAATTCATTTGTCACGAATTGC	518			
Db	425	CAAAAGAGCGACG-TCCTGCTTCCCATGTGCTC-ACCATGCTGGACAACTTCCCAACCA	482			
Qy	519	AGAGTCGGGCATCATATCCCTCATCATCATGTATACAACATTTGATGCTTTACCATGAC	578			
Db	483	TCTACACCCCATGTCTCAGCTGAGTGCAGCTGTTACACGCCCTCAACAGTGAAGTAACTT	542			
Qy	579	AGCTATCAATGACCCAGTTTGGCTACTGGAGTCAIGGCTCTTCAGSTTCAAGCTGAAT	638			
Db	543	TGCCCAAGCATATGACGGGTATCAGCGCGCAACCAAGTACTGGGAGTTGATTTATGAAGA	602			

```

QY 639 TCARAAGCCATACGAGAAGGATTCACAATCAAGATTTGGACCAACATATCAGGA 598
Db 603 TTCTGTGATCTAATAGCAAGCTACCTTGTGTGCGAAGAAATCTACGAAATCTCTA 562
QY 699 TTCCAATGATCTGATGCTCAAGTTCACATTTGCTGCTATGTTTATGCGAGGATGA 758
Db 663 CTGGAAGGCGAGCGGTATTGGGCGCATTTGACTCTAACTGCGACTGCTCTCACAAATTCAC 722
QY 759 CAAGAATGGTGACACTATACCTAAGGATGATC---CCTGGATATGGTGCAATTTTGC 815
Db 723 CAACATGTTAGGCTAATCT-GATCATCAGTTTACTGAGCTC--ATGCGGCTGTACCTCAG 779
QY 816 TCACATGCTGGTTTTCAGTAGCTCTGAATGTCATGAATCTTCTATGAGGCTCTATGTAAC 875
Db 780 CATCCACAGTGCACATGAGGCGCAATGTAAGTGCCCATACACAGCCACTTGTGGGCGAG 839
QY 876 AATACAGAGTATCATGAAGGTGTAATGTCAGTGCACACCGGTCACATTTGTTGCTAG 935
Db 840 TGCCCTTTCCGACCCCTTACCTGTCTCTTTCAGCAGCCATGAAGCGGCTGGCAGGCGCTCT 899
QY 936 TSCITTTGCTGAICCTTACCTCTCTTTTGTGCTGCTTTGAATGGTTTACCGGACCACT 995
Db 900 CCATGAGCTGGCAATACAGGAAGTGTGCTGTGCTTAACA-CAGCTGCGAAGGAAGTTG 958
QY 996 TCATGGTTTACCAATCAGGAAGTTTGTCTATGGATAAAATCTGTTGTAAGAATGTGG 1055
Db 959 GCAAAAGATGTGCAGATGAGAAGTTACGAGACTACATCTGGAACACACTCAACTCAGGAC 1018
QY 1056 GGAGA-ACATTTCCAAAGAGCAGTTGAAGACTATGTTTGGAAACATTTGAACAGTGGCA 1114
Db 1019 GGGTTGTTCCAGGCTATGGCCATGCACTACTAAGGAAGACTGATCCGCGATATACCTGTC 1078
QY 1115 AGGTTGTCCTGGTTTGGACATGGAGTTCTGCGAAAGACTGTACCAAGATATACATGCC 1174
Db 1079 ACGGAGATTTGCTCTGAAACACCTGCTTAATGACCCCATGTTTAACTGGTTGCTCAGC 1138
QY 1175 AGAGAGATTTGCTATGAAGCATTTGCTGGAAGATCCACTGTTTCACTGGTTTCAAAAC 1234
Db 1139 TGTACAAGATTGGCCCAATGTCTCTTAGAGCAGGGTAAAGCAAGAACTCTTGGCCCA 1198
QY 1235 TCTACGAAGTTTCTCC-TGTTCTACAGAACTTGGCAAGTTAA-AA-CTTTGGCCAA 1291
Db 1199 ATGTAGATCTCACAGTGGGTTGCTGCTCCAGTATTATGCGATGACGAGATGAATTA 1258
QY 1292 ATGTTGATGCCACAGTGGTGTGTTGTTGAACATATTATGTTTAACTGAAGCAAGATATT 1351
Db 1259 ACACGGTCTGTTTGGGTTGTCAGGAGATTTGGTGTACTGGCACAGCTCATCTGGAGCC 1318
QY 1352 ATACGGTCTCTTTGGCGTATCAAGAGCTCTTGGCATTTGCTCTCAGCTAAATTTGGGACC 1411
Db 1319 GAGCCTTAGGCTTCCCTCTAGAAAGGCCCAAG 1350
QY 1412 GAGCTCTTGGATTGCGCTAGAGAGGCCCAAG 1443

RESULT 12
LOCUS PIGCITSYN 1455 bp mRNA MAM 17-OCT-1991
DEFINITION Porcine citrate synthase mRNA, complete cds.
ACCESSION M21197
NID g164418
VERSION M21197.1 GI:164418
KEYWORDS citrate synthase.
SOURCE Porcine kidney, cDNA to mRNA, clone PCS4.
ORGANISM Sus scrofa
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Suiformes; Suidae; Sus.
REFERENCE 1 (bases 1 to 1455)
AUTHORS Evans,C.R., Owens,D.D., Sumegi,B., Kispal,G. and Srere,P.A.
TITLE Isolation, nucleotide sequence, and expression of a cDNA encoding
pig citrate synthase
JOURNAL Biochemistry 27, 4680-4686 (1988)
MEDLINE 89000665

```

```

FEATURES
source 1..1455
/organism="Sus scrofa"
/db_xref="taxon:9823"
CDS
7..1401
/translation="MALDTAAARLFGAKNASCLVLAARHASASSTNLKDLIALDLPKE
QAKIKTFROQHNTVVGQITVDMMYGMGRMKGLVYETSVLDPOEGTIFRYSYIPECO
KMLPKAKGEEPLPEGLWLLVGTQIPTEOVSWLSKEWAKRAALPSRVMTJDNFPT
NLHPMSOLSAATALNSENFAVABEGIHRTKYWELIYEDCMDLIAKLPVAAKIYR
NLYREGSSIGADSKLDNSHNTNMLGYTDAQFTLMRLYLTIHSDHEGNNVSAHTYR
LVGSALSDPYLSFAAMNCLAGPLHGLANQEVVLWLTOLQKEVSKDVSDEKLRYIWN
LTNSGRVPGYGHAVLRKIDPRYTCOREFALHKLPHDFMFLVAQLYKIVPNVLEQG
KAKNPWPNVDHAGSVLLQYYGTMENNYTVLFGVSRALGVLAQLIWSRALGPELPERK
SMSTDGLIKLVDSK"
8..88
sig_peptide
/translation="citrate synthase signal peptide"
89..1399
mat_peptide
/translation="citrate synthase"
BASE COUNT 373 a 363 c 386 g 333 t
ORIGIN Chromosome 12.
Query Match 12.7%; Score 241; DB 23; Length 1455;
Best Local Similarity 63.3%; Pred. No. 3.19e-188;
Matches 741; Conservative 0; Mismatches 416; Indels 14; Gaps 13;
Db 193 ATCACTGTGCACATGATGTATGTGTGCATGAGAGGATGAAGGATTTGGTGTATGAACA 252
QY 280 ATCACTGTGATATGTTCTTGTGGAATGAGAGGAATGACAGGATTAATCTGTGAAACCT 339
Db 253 TCGGTTCTGATCCTGATGAGGCACTCCGTTTCCGAGGCTACAGTATCCTCTGAATGCCAG 312
QY 340 CATTACCTTGACCTGATGAGGAATTCGTTCCGGGGTGTGCTATACCTGAATGCCAA 399
Db 313 AAAATGCTGCCAAGGCTAAGGGTGGGAAAGAACCGTCCAGAGGCTTATTTTGGCTG 372
QY 400 AAGGTATTACCTGCAGCAAGCCCTGGGGTGTAGCCCTTGCTCGAAGTCTTCTCTGGCTT 459
Db 373 CTGTTAAGTGCACAGATGCCAACAGAGGACGAGTGTCTTGGCTC-TCAAAAGAGTGGGC 431
QY 460 CTTTAAACAGAAAGGTGCCATCAAAAGAGCAAGTGAATTCATTTCTCAGGAATTCGA 519
Db 432 AAAGAGGCGACG-TCTGCTTCCCATGTGCTC-ACCATGCTGCACACACTTCCACGAA 489
QY 520 GAGTCGGGCAATATATCCCTGATCATCATGTATACAACTATGATGCCITACCAGTCACA 579
Db 490 CTACACCCCATGCTCAGCTCAGTCAGCCATTAACAGCCCTCAACAGTGAAGTAACTTT 549
QY 580 GCTCATCAATGACCCAGTTTGTACTGGAGTCATGCTTCCAGGTTCAGAGTGAATTT 639
Db 550 GCCCGAGCATATGACAGAGGTATCCACCAGAACCAAGTACTGGAGTGTGATTATGAAGAC 609
QY 640 CAAAAGGSCATACGAGAAGGATTCACAAATCAAAAGTATITGGAAACCAACATATAGGAT 699
Db 610 TGTATGATCTGATCGCAAGACTACCTTGTGTGCAAGCAAGATCTACCGGAATCTCTAC 669
QY 700 TCCATGATCTGATGCTCAAGTTCACCTTGTGCTTATGCTTATGCTGAGGATGAC 759
Db 670 CGGGAAGCAGCAGTATTTGGGCAATGATTCTTAAACTGGAGTGTGCCACAAATTTACCC 729
QY 760 AAGAATGGTGACACTATACCTAAGGATGAATCCCTGGATTA-TGGTGC-A-AATTTTGTCT 816
Db 730 AACATGTTAGGCTATATCTGATGCTCAGTT-CAGGGAGCTC-ATGCGCTGTACCTCACC 786
QY 817 CACATGCTTGGTTTTCAGTAGCTCTGAAATGATCATCAACTCTTATGAGGCTCTATGAACA 876
Db 787 ATCCACAGTGCACCATGAAGGTGGCAATGTAAGTGTCTCATACAGCCATTTGGTGGCAGT 846
QY 877 ATACACATGATCATGAAGGTGGTATGTGTCAGTCTCACACCGCTCAGTTGGTGTAGT 936

```

```

Db 847 GCGCTTTCAGACCCCTACTTGTCTTTCGACGAGCCATGAATGGCTGGCAGGCCCTTA 306
QY 937 GCTTTCTGATGCTTACCTCTCTCTTGTCTGCTTTGATGGTTTAGCGGACCACTT 996
Db 907 CATGGGCTGGCAATCAGGAAGTGTCTTGTGGCTGACA-CAGCTGCGAGAAGTGG 965
QY 997 CATGGTTAGCCCAATCAGGAAGTGTCTGATGGATAAATCTGTGTAGAGAATGGGG 1056
Db 966 CAAAGATGTGTACAGATGAGAAGTACGAGACTACATCTGGAATACACTCAACTCAGGACG 1025
QY 1057 GAGA-ACATTTCCAAAGACGAGTGAAGACTATGTTTGGAAACATTTGAACAGTGGCAA 1115
Db 1026 GGTGTCCAGCTATGCGCAGCAGTACTAGGAAGACTGATCACGATATACCTGTCA 1085
QY 1116 GGTGTCCCTGTTTGGACATGGAGTCTCGGAAGACTGTACCAAGATATACATGCCA 1175
Db 1086 AAGAGATTTGTCTGTAACACCTCGCTCATGACGCCATGTTTAACTGTGTGTACGCT 1145
QY 1176 GACAGAGTTGCTATCAAGCATTTGCTGAAGATCCACTGTTCAACTGTGTTCAAACT 1235
Db 1146 GTACAGATTTGCCCAATGCTCTCTGGAACAGGCGAAGCTAAGATCTCTGGCCCAA 1205
QY 1236 CTACGAAGTTTCCCTCC-TGTTCTTACAGAACTTGGCAAGTTAA-AA-CCTTGGCCAAA 1292
Db 1206 TGTGGATGCTCAGCTGGGCTGCTCTCCAGTACTATGCGATGCGGAGATGAATCTACTA 1265
QY 1293 TGTGTAGCCCAAGCTGCTGTTGTGTAATCTATGTTTAACTGAAGCAAGATATTA 1352
Db 1266 CACAGTCTGTTGGGGTATACAGGCGCACTGGGTGTTATAGCACAGCTCATCTGGAGCCG 1325
QY 1353 TAGGGTCTCTTTGGCGTATCAAGAGCTCTTGGCATTTGCTCTCAGCTAAATTTGGGACCG 1412
Db 1326 AGCCTTAGGCTTCCCTCTAGAGAGCCCAAG 1356
QY 1413 AGCTCTGGATTGCGCTAGAGAGCCCAAG 1443

RESULT 13
LOCUS SCMTCSINA 1481 bp DNA PLN 17-FEB-1997
DEFINITION S.cerevisiae mitochondrial citrate synthase gene, complete CDS.
ACCESSION Z23259
NID 9313749
VERSION Z23259.1 GI:313749
KEYWORDS mitochondrial citrate synthase.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 1481)
AUTHORS Lindner,P. and Pluueckthun,A.
JOURNAL Biochemie
REFERENCE 2 (bases 1 to 1481)
AUTHORS Lindner,P. and Pluueckthun,A.
JOURNAL The effect of point mutations in the hinge of yeast citrate synthase
REFERENCE 3 (bases 1 to 1481)
AUTHORS Lindner,P.
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 1481)
AUTHORS Lindner,P.
JOURNAL Direct Submission
REFERENCE 5 (bases 1 to 1481)
AUTHORS Lindner,P. and Pluueckthun,A.
JOURNAL Submitted (05-JUL-1993) Peter Lindner, Protein Engineering, Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a, Martinsried, D-82152, Fed. Rep., Germany
REFERENCE 6 (bases 1 to 1481)
AUTHORS Suissa,M., Suda,K. and Schatz,G.
JOURNAL Isolation of the nuclear yeast genes for citrate synthase and fifteen other mitochondrial proteins by a new screening method
MEDLINE EMBO J. 3 (1984) 1773-1781 (1984)
REMARK 8503587
FEATURES (sites) Location/Qualifiers
SOURCE 1. .1481

```

```

/organism="Saccharomyces cerevisiae"
/strain="D273-108"
/db_xref="taxon:4932"
/clone="plasmid pFCS1 (Suissa et al., EMBO J. 3, 1773-1781 (1984))."
1. .1481
/gene="CIT1"
/db_xref="SGD:L0000341"
1. .111
/gene="CIT1"
/db_xref="SGD:L0000341"
/evidence="experimental"
/product="the processing site has been determined by direct N-terminal protein sequencing"
1. .1440
/gene="CIT1"
/codon_start=1
/db_xref="SGD:L0000341"
/product="mitochondrial citrate synthase"
/protein_id="CAA80781.1"
/db_xref="PID:g313750"
/db_xref="GI:313750"
/db_xref="SWISS-PROT:P00890"
/translation="MSAIISTTSKLSRSTQCONMOKALFALLNARHYSASEQT
LKERFABIIPAKAEIKKKEKHTVIGEVLEQAYGGMRIKGLWEGSVLDPEEG
IRFERGTIPRIQELPKAGSTPLPALFWLLTGEITDAQVKALSADLARGSEIP
EHVGLDLSLPKOLHPMAQPSIAVTALESEKFAKAYAGVSKKEYSVTFEDSLDL
GKLVIAKSIYRNVDKGTISTDPNDYKGNLAQLLGENKDFIDLMRLYLTLHSDH
EGGNSAHTHLVGSALSSPYLSAAGNLGAPLHGRANQEVLEWFLKLRVEKGDY
SKETIERYLMDTLNAGRVPGYHVLKRTDPRTAOREFALKHFPDLPKLVSTIY
EVAPGLTKHGTKNPMPNDVSHGVLQVYGLTEASFYTVLFGVARAIGVLPQLIID
RAVCAPIERPKFSTEKYKELVKKIESKN"
172
/gene="CIT1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace="g"
234^235
/gene="CIT1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace=""
1472
/gene="CIT1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace="a"
452 a 266 c 319 g 444 t

BASE COUNT 452 a 266 c 319 g 444 t
ORIGIN
Query Match 10.2%; Score 193; DB 27; Length 1481;
Best Local Similarity 61.2%; Pred. No. 1.82e-143;
Matches 703; Conservative 0; Mismatches 432; Indels 13; Gaps 12;
Db 243 TGGTGTATGAGAGGTATTAAGGCCCTGTTTGGGAAGGTTCCCTGTTAGACCCCGAAGA 402
QY 300 TGGTGAATGAGAGGATGACAGGATTACTGTGGAACCTCAITACCTGACCCGTGATGA 359
Db 303 AGGTATTAGATTAGGGCTGCTACTATTCCAGAAATTCAGAGGCACTACCAAGGCTGA 362
QY 360 GGAATTCCTCCGGGGTGTCTATACCTGAATGCCAAAGGATATTACTGCGAGCAA 419
Db 363 GGGTAGTACAGAACCTTTGCCAGAGCTTTATTTGGTTGCTTTTGTAGTGTGAATACC 422
QY 420 GCTTGGGGGTGAGCCCTTGCCTGAAGSTCTTCTCTGGCTTTTAAAGCAAGAGTGGC 474
Db 423 TACTGAGCTCAAGTTAAGGCCCTT-ICTGCTGATTAGCTGCCAGATCAGAA-ATTCGA 480
QY 480 ATCAAAGAGCAAGTGAATTCATTTGCTCAGSAAATTCAGTGGGCAATCATATCCTT 534
Db 481 GAGCA-CGT-TATCCAACTTTTAGATAGCCCTCCCAAGAGATCTATATCCAAATGGGCAAT 538
QY 540 GATCATCATGATACAACTATT-GATGCCCTTACCAAGTTCACAGCTCATCATCAAGCCAGT 598

```

Db	539	TTTCTATTGCGGTACTGCTTTTGAAGACGAGTCTAAAGTTTGCACCAAGCATATGCTCAAG	599
Qy	599	TTGCTACTGGAGTCAAGCTCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	658
Db	599	GTGTATCCCAAGAAAGATATTGGAGCTATACATTTGAAGATTGCTAGATCTGCTGGTA	658
Qy	659	GGATTCAAAATCAAGATTGGGAACCAACATATGAGGATTCATGAATCTGATTGCTC	718
Db	659	AATTACCTGTTATGCTTCCAAATTTATCGTAATGTTTCAAGATGGTAAATATTACTT	718
Qy	719	AGATTCCACTTTGCTGCTTATGTTTATCCAGAGTGTACAAAGATGGTGACACTATAC	778
Db	719	CAACCGATCCCTAATGCTGACTATGTAATAAATTTGGCCCAACTTTTGGGCTAGCAAAACA	778
Qy	779	CTAAGGATGAATCCCTGATGATGTCGCAAAATTTTGTCTACATGCTTGGTTTCAAGTCT	838
Db	779	AGGATTTTATTGA-CTTA--ATGAGACTATATTAACTATTCTGATCATGAAGGTG	835
Qy	839	CTGAATGCAATGAATCTTATGAGGCTCTATGTAACAATACACAGTATCATGAAGGTG	898
Db	836	GTAGGTTTCTGCCATACATACATTTAGTGGTCTGCTTATCTATCTGCGCATCTAT	895
Qy	899	GTATGTCAGTCTCACACCGCTACTTGGTGTAGTGTCTTGTCTGATCTTCACTCT	958
Db	896	CTTTGGCGCTGGTTTGAATGGTTTAGTGGCCCATACATGCTGCTGCCAATCAAGAAG	955
Qy	959	CTTTGCTGCTCTTGAATGGTTTAGCGGACCACTTCATGTTTAGCCCAATCAGGAAG	1013
Db	956	TTTTAGATGGCTATTAAATTTAGAGAAAGTGAAGGTGACTATTTCAAAAGAAACAA	1015
Qy	1019	TTTGTCTATGGAATAAATCTGTTGAGAAGAA-TCTGGGGAGAACATTTCCAAAGAGCAG	1077
Db	1016	TTGAAA-AGTACTTGTGGATACITTTGAACGACGAGAGTGTCTGCTGTTATGGCCAT	1074
Qy	1078	TTGAAAGACTATGTTTGAAGAAACATTTGAACAGTGGCAAGGTGTCTGCTGTTTGGACAT	1137
Db	1075	GCAGTTTGTGAGAAACATGATCCTGTTTACGGCTCAAGGTGAATTCGCAATTTGAACAT	1134
Qy	1138	CGAGTTCTGGAACACATGTACCAAGATATACATGCCAGAGAGAGTTCGCTATGAAGCAT	1197
Db	1135	TTCCAGATTAGGATTTAATGTTGCTCTCCACATTTATGAAGTGGCCCGAGGGTT	1194
Qy	1198	TTGCTGGAAGATCCACTGTTTCAACTGCTTTCAAACTCTACGAAGTTTTCCTCCCTGTC	1257
Db	1195	TTAATTAAGCATGTAAACTAAGAACCCATGGCCAAATGTTGATTTCACATTCGCGGT	1254
Qy	1258	TTACAGAA-CTTGGCAAGTTAA-ACCT-TGGCCAAATGTTGATGCCACAGTGGTG	1314
Db	1255	TTATTGCAATATGTTCTAAGTGGCTTCTGCTTCTACAGTGTATGTTTGGTGTGTC	1314
Qy	1315	TTGTTGCACTATTATGTTTAACTGAAGCAAGATATTATAGGTCCTCTTGGCGTATCA	1374
Db	1315	AGAGCTATGTTGCTTACCCCAATTTATCATCATAGGCTGTTGGTCTCCCAATCGAA	1374
Qy	1375	AGAGCTCTTGGCAATTTGCTCTCAGCTAAATTTGGGACCGAGCTCTTGGATTGGCGCTAGAG	1434
Db	1375	AGGCCAAA	1382
Qy	1435	AGGCCAAA	1442
RESULT 14			
LOCUS Yeast gene for citrate synthase. 2427 bp DNA			
DEFINITION Yeast gene for citrate synthase.			
ACCESSION X00782			
NID Q3602			
VERSION X00782.1 GI:3602			
KEYWORDS citrate synthase; synthase.			
SOURCE baker's yeast.			
ORGANISM Saccharomyces cerevisiae			
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;			
Saccharomycetaceae; Saccharomyces.			

REFERENCE	1 (bases 1 to 2427)
AUTHORS	Suissa,M., Suda,K. and Schatz,G.
TITLE	Isolation of the nuclear yeast genes for citrate synthase and fifteen other mitochondrial proteins by a new screening method
JOURNAL	EMBO J. 3 (8), 1773-1781 (1984)
MEDLINE	85003587
COMMENT	Data kindly reviewed (09-MAY-1985) by G. Schatz.
FEATURES	Location/Qualifiers
source	1..2427
	/organism="Saccharomyces cerevisiae"
	/db_xref="taxon:4932"
CDS	376..1818
	/note="citrate synthase"
	/codon_start=1
	/protein_id="CAA25359.1"
	/db_xref="PID:g3603"
	/db_xref="GI:3603"
	/db_xref="SWISS-PROT:P00890"
	/translation="MSAILSTSKSFLSRGSTRQCNQMKALLNARHYSSASEQT LKRFAPETPAKAEIKKFKHGTIVGEVLEBQAYGGMRIKGLVWEGSVLDPEE GIRFGRTIPEIQRELPAKAGSTELPEALFWLLTGTIPDAQVKALSAIDARSEL PEHVILDSLPDLHPMAQFSIAVTALESSEKPAKAYAGVSKSEYFSDLSL LGLPVIASKITRNPKDKIITIDNADYGRNLAQLLGYNKDFIDLMRLIILHSD HGGNYSATHTHLVGSALSPYLSLAAGNLGAPHRANOEVLEWLFKREEVKG YSKETIEKYLDITLNAGRVPGYHAGLVKRTDPRYTAQREFAKHFPDYELFKVSTI YEVAPGLTKHGKTKNPWNVDSHGVLLQVYGLTEASFTVLFVARAIGVLPOLII DRVAGPIERPKSFSTEKYKELVKIESKN"
BASE COUNT	757 a 439 c 474 g 757 t
ORIGIN	
Query Match	10.2%; Score 193; DB 27; Length 2427;
Best Local Similarity	61.2%; Pred. No. 1.82e-143;
Matches	703; Conservative 0; Mismatches 432; Indels 13; Gaps 12;
Db	621 TGGTGGTATGAGAGGTATTAAAGGCTTGTGTTGGGAAGGTTCGCTGTAGACCCGGAAG 680
Qy	300 TGGTGGTATGAGAGGTATTAAAGGCTTGTGTTGGGAAGGTTCGCTGTAGACCCGGAAG 359
Db	681 AGGTATTAGATTTAGGGGTGCTACTATTCCAGAAATTCAAAGGGAACCTACCAAGGCTCA 740
Qy	360 GGAATTCGCTTCGGGGGTGCTATACCTGAATGCCAAGAGTATTACCTGCACGAAA 419
Db	741 GGGTAGTACAGAACCTTTGCCAGAGCTTTATTGTTGGTTCGCTTTTACGTGGTGAATAAC 800
Qy	420 GCTTGGGGGTGAGCCCTTGCCTGAAGGTCTCTCTGCTGCTCTTTTAAAGGAAAGGTGCC 479
Db	801 TACTGACGCTCAAGTTAAAGCCCTT-TCTGCTGATTAGTGCACAGATCAGAA-ATTCCA 858
Qy	480 ATCAAAAGAGCAAGTGAATTCGAATTTGCTCAGGAATTCAGAGTTCGGGCATCATATCCCT 539
Db	859 GAGCA-CGT-TATCCAACTTTTAGATAGCTCCCAAGATCTACATCCAAATGGCGCAAT 916
Qy	540 GATCATGATATACAACTATT-GATGCCCTACCAAGTTCACAGCTCATCCAATGACCCAGT 598
Db	917 TTTCTATTGCGGTGACTGCTTTAGAAGCGAGTCTAAGTTTGCAGAGCATATGCTCAAG 976
Qy	599 TTGCTACTGGAGTCATGGCTCTTCAAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGT 658
Db	977 GTGTATCCAGAAAGATATTGGAGCTATACATTTGAAGATTCGTTAGATCTCTGGGTA 1036
Qy	659 GGATTCAGAAATCAAGTATTGGGAACCAACATATGAGGATTCATGATCTGATTGCTC 718
Db	1037 AATTACCTGTTTATGCTTCCAAATTTTATCGTAATGTTTCAAGGATGGTAAATTTACTT 1096
Qy	719 AAGTTCACCTTTTGTGCTGCTTATGTTTATCGAGGATGTACAAGAATGGTGACACTATAC 778
Db	1097 CAACCGATCCTTAATGCTGACTATGGTAAAAATTTGGCCCAACTTTTGGGCTACGAAAAACA 1156
Qy	779 CTAAGGATGAATCCCTGGATTTATGGTCAAAATTTTGCACATGCTTGGTTTCAGIAGCT 838
Db	1157 AGGATTTTATTGA-CTTA--ATGAGACTATATTAACTATTCTGATCATGAAGGTG 1213
Qy	839 CTGAATGCATGAATCTCTTATGAGGCTCTATGTAACAATACACAGTATCATCAAGGTG 898

```

Db 1214 GTAACGTTTCTGCCATACTACACATTTAGTGGGTTCTGCCTTATCTCTCGCCATCTTAT 1273
QY 899 GTAATGTCAGTGTCTACACGCGTCACTTGGTTGCTAGTGGCTTGTCTGATCTTACCTCT 958
Db 1274 CTTTGGCGGTGTTTGAATGTTTAGTGGCCCATTTACATGGTGTGTCGCCAATCAAGAAG 1333
QY 959 CTTTGGCTGCTTGTGAATGTTTAGTGGCCCATTTACATGGTGTGTCGCCAATCAGGAAG 1018
Db 1334 TTTTGAATGGGTATTTAAATTTGAGAGAGAGTGAAGGTTGACTATTTCAAAAGAAACAA 1393
QY 1019 TTTTGTATGGATAAAATCTGTTGTAGAAGAA-TGTGGGGAGAACATTTCCAAAGAGCAG 1077
Db 1394 TTGAAA-AGTACTCTGCGGATCTTTGACCCAGGAGAGTGTTCCTGGTTATGGCCAT 1452
QY 1078 TTTGAAAGACTATGTTTGGAAAACATTTGAACAGTGGCAAGGTTTGGCTTGTGGACAT 1137
Db 1453 GCGGTTTGTGAGAAAACATGCTCTGTTATAGCGGTCAACGTGAATTCGCATTTGAAACAT 1512
QY 1138 GGAGTTCTCGGAAGACTGTACCAAGATATACATGCCAGAGAGAGTTCGCTATGAAGCAT 1197
Db 1513 TTCCAGATTACAGTTATTTAAGTTGGTCTCCACCATTTATGAAGTTTGGCCAGGGGTT 1572
QY 1198 TTCCCTGAAGATCCCATGTTTCAACTGGTTCCTCAAACTCTACGAAGTTTTCCTCTCTTC 1257
Db 1573 TTAACCTAAGCATGGTAAACTTAAGAACCCATGGCCAAATGTTGATTCACATTCCTCGGT 1632
QY 1258 TTACAGAA-CTTGGCAAGATTAACCT-TGGCCAAATGTTGATGGCCACAGTGTGTG 1314
Db 1633 TTATGCAATACATAGTCTAACTAGGCTTGGTTCTACACTGATTTGTTGGTGTGCC 1692
QY 1315 TTGTTGAACATATATGTTTAACTGAAGCAAGATATTACGGTCTCTTTGGCGCTATCA 1374
Db 1693 AGAGCTATTGGTGTGTACCCCAATTAATCATCGATAGGCTGTGTGTCCTCAATGAA 1752
QY 1375 AGAGCTCTTGGCATTTGCTCTCAGCTAATTTGGGACCGAGCTCTTGATTCGCCCTAGAG 1434
Db 1753 AGGCCAAA 1760
QY 1435 AGGCCAAA 1442

RESULT 15
LOCUS SCYNR001C 4391 bp DNA PLN 11-AUG-1997
DEFINITION S cerevisiae chromosome XIV reading frame ORF YNR001c.
ACCESSION 271616 Y13139
NID g1302468
VERSION 271616.1 GI:1302468
KEYWORDS
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 4391)
AUTHORS Aert,R., Verhasselt,P., Voet,M. and Voelckaert,G.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4391)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
source
1..4391
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XIV"
misc_feature 142..251
misc_feature 142..149
misc_feature 150..234

```

```

/misc_feature
235..259
/note="CEN element CDE II"
/gene
complement(1006..2445)
/db_xref="SGD:L0000341"
complement(1006..2445)
/feature="CDE II"
/feature="ORF YNR001c"
/codon_start=1
/db_xref="SGD:L0000341"
/protein_id="CAA96277.1"
/db_xref="PID:e239778"
/db_xref="PID:g1302469"
/db_xref="GI:1302469"
/db_xref="SWISS-PROT:P00890"
/translation="MSAIISTTSKLSFGSTQCONMOKALFALLNARHYSASEQT
LKERFATIIKPAKEIKKFKKHGKTVIGEVLEQAYGGMRGIKGLVWEGSVLDPEEG
IRFRGRTIPIORLPAKESGTEPLPEALFWLLLTGEIPTDAQVKAISADLAARSEIP
EHVLOLDSLPKDLHPMAOFSAIVTALESSEKFAKAYAGVSKKEYWSYTFEESLDLL
GKLPVASKIYRNVFQDKITSDPNADYKKNLAQLLGTENKDFIDLMRLYLTTHSDH
EGNVSAHTTHLVGSLSSPYLSAAGLNGLAGPLHGRANQEVLEWFLKREEVKSTY
SKETIERKLDITLNAGRVVPGYGHAVLRKTDPRYTAQREFALKHFFDFELKLVSTIY
EAPGVLTGKTKNPNVNDSHSGVLLQYVGLTEASFYTVLFGVARAIGVLPQLIID
RAVGAPIERPKSFSTKIKELVKKIESKN"
complement(3230..3301)
/feature="tp(AGG)NR - systematic name"
/feature="anticodon gene: AGG; tRNA-Pro2 - common name"
complement(3230..3301)
/feature="tp(AGG)NR - systematic name"
complement(3382..3497)
/feature="delta-remnant"
complement(3503..3873)
/feature="solo-tau"
3983..4056
/feature="tn(GTT)NR - systematic name"
/feature="anticodon gene: GTT; tRNA-Asu - common name"
3983..4056
/feature="tn(GTT)NR - systematic name"
BASE COUNT 1412 a 803 c 784 g 1392 t
ORIGIN

Query Match 10.2%; Score 193; DB 27; Length 4391;
Best Local Similarity 61.2%; Pred. No. 1.82e-143;
Matches 703; Conservative 0; Mismatches 432; Indels 13; Gaps 12;

Db 1064 TTTGGCCTTTTCATTTGGAGCACCACAGCCCTATCGATGATTAAATTTGGGGTAACACCA 1123
Cp 1442 TTTGGCCTCTCTAGCGCAATCCCAAGAGCTGGTCCCAAAATAGCTGAGACCAATGCCA 1383
Db 1124 ATAGCTCTGGCAACACCAACAATACAGTGTAGAACGAAGCTCAGTTAGACCATAGTAT 1183
Cp 1382 AGAGCTCTTGATACCCCAAGACAGCCGTATATATCTGCTTCAGTTAAACCATATATC 1323
Db 1184 TCGAATAAACACCGGAATGTGAATCAACATTTGGCCATGGGTTCTTAGTTTACCAATGC 1243
Cp 1322 TTCAACAACACACACCACTCTGGCCATCAACATTTGGCCAAAGG-TT-TTAACTTTGCCAAGT 1265
Db 1244 TTAGTTAAACCCCTGGGCAACTTCATAATGTGGAGACCAACTTAAATACTCTAA 1303
Cp 1264 TCTGTAAAGACAGGAGAA-ACTTCGTAGAGTTTGAACCCAGTTTGAACAGTGGATCT 1206
Db 1304 TCTGGAAATGTTTCAATGCGAATTCAGTTGAGCCGTATACAGGAGATCAGTTTTCIC 1363
Cp 1205 TCAGGCAAAATGTTTCATAGCGAACTCTCTCGGCATGTATATCTTGTACAGCTTTTCG 1146
Db 1364 AAAACCGCATGCCATACAGGAAACACTCTCCCTGGTTCAAGTATCCCAAGTAC 1423
Cp 1145 AGAAGTCCATGTCCAAACACAGGAGCAACCTTGGCACTGTTCATATGTTTCCAAACATG 1086
Db 1424 T-TTCAATTTGTTCTTTTGAATAGTACCTTTCACCTTCTCTCTCAATTTAAATACCA 1482
Cp 1085 TCTTTCAACTGCTCTCTTGGAAATGTTCTCCCCACA-TTCTTCTACACACATTTTATCCA 1027

```


Db 1483 TTCTAAACCTCTTGATTGGCAGCAGCATGTAAATGGCCAGCTAAACCATTCACAAACGAGC 1542
 Cp 1026 TAGCAAACTTCCTGATTTGGCTAAACCATGAAGTGGTCCGGCTAAACCATTCACAAACGAGC 967
 Db 1543 GGCCTAAAGATAGTATGGCAGGATAGGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1602
 Cp 966 AGCAAGGAGAGGTAAAGATCAGACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 907
 Db 1603 AACGTTACCACTTCATCATCAGATGAATAGTAAATATAGTCTCATTA--AGTCAATA 1660
 Cp 906 GACATTACCACTTCATCATCAGTGTGATTTGTACATAGAGCCTCATAGAAGTTCAATG 847
 Db 1661 AAATCCCTTGTCTTCCATGCCCAAAAGT--TGGGCCAAATTTTACCATAGTACGCAATTAGG 1719
 Cp 846 CATTTACAGAGCTACTGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 787
 Db 1720 ATCGGTTGAAGTAATTTTACCATCCTTGAACACATTTAGCATAAATTTTGAAGCAATAAC 1779
 Cp 786 ATCCCTTAGGTATAGTGTCAACCAATTTCTGTACATCCTCGATAAACATAGCAGCAACAAG 727
 Db 1780 AGGTAATTTACCCAGCAGATCTAACGAATCTTCAATGTATAGCTCCAATATTTCTTT 1839
 Cp 726 TCGAATTCAGCAATCAGATTCATGGAATCCTCATATGTTGGTTCCCATACTTTGATTT 567
 Db 1840 GGATACACCTTGAGCAATATGCTTTGGCAAACTTAGACTCGCTTTCTAAAGCAGTCAAGGC 1899
 Cp 666 GTGAATCCCTTTCTCGTATGCCCTTTGAAATTCATTTGAACCTGAAGAGCCATGACTCC 607
 Db 1900 AATAGAAATTCGCCCATTTGGATAGATCTTTTGGAGGCTATCTAAAAGTTGGATA-A 1958
 Cp 606 AGTAGCAAACTGGGTTCATTTGGATGAGCTGTGACTGGTGAAGGCATC-AATAGTTGTATACA 548
 Db 1959 CG-TGCTCTGGAATTTCTGATCTGGCAGCTAAAT-CAGCAGA-AAGGGCTTTAACTTGAG 2015
 Cp 547 TCATGATCAGGATATGATGCCGACCTGCAATTCCTGAGACAAATTTGAATTCATCTGCT 488
 Db 2016 CGTCAGTAGGTATTTTCAACAGTCAAAAGCAACCAAAATAAGGTTCTGGCAAGGTTCTG 2075
 Cp 487 CTITTTGATGGCACCTTTCTGTTAAAGAGAGCCAGAGAACCTTCAGGCAAGGGGCTCAC 428
 Db 2076 TACTACCTTCAGCCCTTTTGGTATGTTCCCTTTTGAATTTCTGGAATAGTACGACCCCTAAATC 2135
 Cp 427 CCCAGGCTTTGCTCGAGGTAAATACCTTTTGGCAATTCAGGTATAGACACCCCGGAAAGC 368
 Db 2136 TAATACCTTCTTGGGGTCTTAACACGGAACTTCCCAACAGGCTTTTAAATACCTCTCA 2195
 Cp 367 GAATTCCTTCATCAGGCTCAAGGTAATGAGGTTTCCACAGTAATCTCTGTCATCTCTCA 308
 Db 2195 TACCACCA 2203
 Cp 307 TTCCACCA 300

Search completed: Sat Oct 23 15:30:05 1999
 Job time : 3443 secs.

CC activity the DNA sequences can also be used to identify similar
 CC sequences in the genomes of other plants and in the production of
 CC transgenic plants with altered CS activity.
 SQ Sequence 1891 BP; 512 A; 370 C; 584 T;

Query Match 100.0%; Score 1891; DB 17; Length 1891;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 1891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ttttctgctccatcagcctactgagatgtattccctactggtataaagttaattttttga 60
 Qy 1 ttttttccgttccatcagcctactgagatgtattccctactggtataaagttaattttttga 60

Db 61 ttttccgaagaatgtgttccactgaggtttcgttgcgttccaaagctccctctcca 120
 Qy 61 ttttttccgaagaatgtgttccactgaggtttcgttgcgttccaaagctccctctcca 120

Db 121 gcggtcccaacagtcacaaattgttagcaattctgctgctgggttccaaagctccaaacctctcc 180
 Qy 121 gcggtcccaacagtcacaaattgttagcaattctgctgctgggttccaaagctccaaacctctcc 180

Db 181 gctttgatctgcgtctgagctggttacaaagaattgatttccctaaacacaggtatgccttg 240
 Qy 181 gctttgatctgcgtctgagctggttacaaagaattgatttccctaaacacaggtatgccttg 240

Db 241 aaaaagatcaactcagatataaagggttcaatttgggaacatcacagtgtgatgtgtctt 300
 Qy 241 aaaaagatcaactcagatataaagggttcaatttgggaacatcacagtgtgatgtgtctt 300

Db 301 ggtgaattgaggaatgacaggattactgtggaaacctcattaccttgacctgtatgag 360
 Qy 301 ggtgaattgaggaatgacaggattactgtggaaacctcattaccttgacctgtatgag 360

Db 361 ggaattcgtctccgggggtgtctataccctgaatgcacaaaggtattacctgcagcaag 420
 Qy 361 ggaattcgtctccgggggtgtgtctataccctgaatgcacaaaggtattacctgcagcaag 420

Db 421 cctgggggtgagcccttgcctgaaggcttctctgtggttctttaaagcaggaaggtgcga 480
 Qy 421 cctgggggtgagcccttgcctgaaggcttctctgtggttctttaaagcaggaaggtgcga 480

Db 481 tcaaaagagcaatgaaatcgaatgtctcaggaattcagagtcgggcatcatccctc 540
 Qy 481 tcaaaagagcaatgaaatcgaatgtctcaggaattcagagtcgggcatcatccctc 540

Db 541 atccatcgtatcaaacattgagctctaccagtcacagctcatccaatgacccagttt 600
 Qy 541 atccatcgtatcaaacattgagctctaccagtcacagctcatccaatgacccagttt 600

Db 601 gctactggagtcagctgtcttcagggttcaaaagtgaatttcaaaagcgcatacagagaagg 560
 Qy 601 gctactggagtcagctgtcttcagggttcaaaagtgaatttcaaaagcgcatacagagaagg 560

Db 661 attcaaaatcaaatgttgggaacacacataatgagattccatgaatcgtatgtctca 720
 Qy 661 attcaaaatcaaatgttgggaacacacataatgagattccatgaatcgtatgtctca 720

Db 721 gttccactgtgtgcttattgtttatcgcaggaatgtacaaagatggtgacacataacct 780
 Qy 721 gttccactgtgtgcttattgtttatcgcaggaatgtacaaagatggtgacacataacct 780

Db 781 aaggatgaatccctggatattggtgcaaaatttgcctcacaatgcttgccttccagtaacctt 840
 Qy 781 aaggatgaatccctggatattggtgcaaaatttgcctcacaatgcttgccttccagtaacctt 840

Db 841 gaaatgcatgaactcttataagctctatgtaacacatacacagtcatacgaagtgt 900
 Qy 841 gaaatgcatgaactcttataagctctatgtaacacatacacagtcatacgaagtgt 900

Db 901 aatgctcagtcacacccggctcacttgggtgctagtgttctgtctatccctaccctctcc 960
 Qy 901 aatgctcagtcacacccggctcacttgggtgctagtgttctgtctatccctaccctctcc 960

Db 961 ttttctactgctttgaaatggttttagcggcagcaacttcactggttttagccaatcaggaagtt 1020
 Qy 961 ttttctactgctttgaaatggttttagcggcagcaacttcactggttttagccaatcaggaagtt 1020

Db 1021 ttgctatggataaaaatcgtgttagaagaatgtggggagaaacatttccaaagagcagttg 1080
 Qy 1021 ttgctatggataaaaatcgtgttagaagaatgtggggagaaacatttccaaagagcagttg 1080

Db 1081 aaaaactatgtttgggaacacattgaaacagtgggcaagttgtccctggttttggacatgaa 1140
 Qy 1081 aaaaactatgtttgggaacacattgaaacagtgggcaagttgtccctggttttggacatgaa 1140

Db 1141 gttctgcgaagacgtaccacaatatacatgccagagaaattcactatgaagcatttg 1200
 Qy 1141 gttctgcgaagacgtaccacaatatacatgccagagaaattcactatgaagcatttg 1200

Db 1201 cctgaagatccactgttttcaactgttttcaaaactctcagaagtttttccctctcttctta 1260
 Qy 1201 cctgaagatccactgttttcaactgttttcaaaactctcagaagtttttccctctcttctta 1260

Db 1261 cagaacttgccaaggttaaaaccttggccaactgttgatgcccaagtggtgtgtgttg 1320
 Qy 1261 cagaacttgccaaggttaaaaccttggccaactgttgatgcccaagtggtgtgtgttg 1320

Db 1321 aactattatggttttaactgaagcaaatattatcacgctcctctttgagctatcaagact 1380
 Qy 1321 aactattatggttttaactgaagcaaatattatcacgctcctctttgagctatcaagact 1380

Db 1381 cttggcatttgcctcagcttaatttgggacggagctcttggatgccgctagagagcca 1440
 Qy 1381 cttggcatttgcctcagcttaatttgggacggagctcttggatgccgctagagagcca 1440

Db 1441 aagagtcacaatgaagtggttgagaaccadtgcaagaagacatgaattgtttgaaat 1500
 Qy 1441 aagagtcacaatgaagtggttgagaaccadtgcaagaagacatgaattgtttgaaat 1500

Db 1501 ctccgagcatalaaacacaaatgtataatctctatgaataattgcttgacaaagcactct 1560
 Qy 1501 ctccgagcatalaaacacaaatgtataatctctatgaataattgcttgacaaagcactct 1560

Db 1561 ttcttggggacagaataggtcggcccttcaatgggttaacgaacttcagttcaaaacttc 1620
 Qy 1561 ttcttggggacagaataggtcggcccttcaatgggttaacgaacttcagttcaaaacttc 1620

Db 1621 actgaattgttgaattgtatggtttctcgagactgtcctgaaatttgaacttagtct 1680
 Qy 1621 actgaattgttgaattgtatggtttctcgagactgtcctgaaatttgaacttagtct 1680

Db 1681 agtgattcatttttcttccgaatttccacacgctgacacagctgacacagatgtaaaatta 1740
 Qy 1681 agtgattcatttttcttccgaatttccacacgctgacacagctgacacagatgtaaaatta 1740

Db 1741 ataagtcacatgaatttaactgcgttcttgggttgcacattagacttctgaatgacttccitt 1800
 Qy 1741 ataagtcacatgaatttaactgcgttcttgggttgcacattagacttctgaatgacttccitt 1800

Db 1801 gctgcgaagttagtaactcggctgattcagcgaataaactgaacttctgtagttctctaaa 1860
 Qy 1801 gctgcgaagttagtaactcggctgattcagcgaataaactgaacttctgtagttctctaaa 1860

Db 1861 tttgctaatcttatttgcattgatattatgaa 1891
 Qy 1861 tttgctaatcttatttgcattgatattatgaa 1891

RESULT 2
 ID T04199 standard; cDNA to mRNA; 1891 BP.
 AC T04199;
 DT 25-JAN-1996 (first entry)
 DE Potato citrate synthase cDNA.
 KW Citrate synthase; flower formation; tuber storage; ss.
 OS Solanum tuberosum.
 FH Key Location/Qualifiers

FT cds 73..1485
 PN /tag- a
 PD WO9524487-A.
 PF 14-SEP-1995.
 PR 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WFI: 95-328278/42.
 DR P-PSDB: R82838.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Claim 5: Page 53-56: 87pp: English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of
 CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. a cDNA library was prepd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 1891 BP; 512 A; 370 C; 425 G; 584 T;

Query Match 100.0%; Score 1891; DB 16; Length 1891;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 1891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ttttctccatcagctactgagatgattcccaactggtaaaagtaatttttga 60
 Qy 1 TTTTTCGTCATCAGCCTACTTGAGATGTTATCCACCTGGTAAAGTTAATTTTTGA 60

Db 61 ttttcgcaagaatggtttctaccctagcgtttgcttctgctcgaagctccgctcga 120
 Qy 61 TTTTCGCGAGCAATGGTGTCTACCGTAGCGTTGCTGCTCAAGCTCCGCTCTCGA 120

Db 121 qcggtcccaacagtcacaaatgtagcaattctgctgctgcttcaagtcacaaactctcc 180
 Qy 121 GCGGTCCACAGTCACAAATGTTAGCAATTCGTGCGCTGAGTTCAGTCCAAACCTCTCC 180

Db 181 ggtctgattcgcgtttctgagctggtgacaaagattgattctcgaacacagatgcctg 240
 Qy 181 GGTCTTGATCTGTGAGCTGTGAGCTGTGACAAAGATTGATTCCTGAACACAGATCGCCTG 240

Db 241 aaaaagatcaagtcagatatgaaggttcaattgggaacatcacagttcattgattctt 300
 Qy 241 AAAAAGATCAAGTCAGATATGAAGGTTCATTTGGGAACATCACAGTTGATAGGTTCTT 300

Db 301 ggtggaatgaggaatgacaggaattactgtggaacacctcattacctgacctgag 360
 Qy 301 GGTGGAATGAGGAATGACAGGATTACTGTGGAACCTCATTACCTTGACCCCTGATGAG 360

Db 361 ggaattcgttcggtggttctctacacgtgaatgcacaaagattaccctgcacaaag 420
 Qy 361 GGAATTCGCTTCGGGGGTTGCTCTATACCTGAAATGCAAAAGGATTACCTCGACAAAG 420

Db 421 cctggggtgagccctgctgaaggtctctctgctgctcttcttaacaggaaggtgcca 480
 Qy 421 CTTGGGGGTGAGCCCTGCTGAGGTCTCTCTGCTGCTCTCTTTTACAGGAAGGTGGCA 480

Db 481 tcaaaagagaagtgaattcaattgtctcaggaattgcagatcggtgcatcatccctg 540
 Qy 481 TCAAAAGAGCAAGTGAATTCATTTGCTCAGGAATTCAGAGTCTGGGATCATATCCCTG 540

Db 541 atcaatcatatacaactattgagcttaccagtcacagctcaccatgaaccagttt 600
 Qy 541 ATCATCATGATACAACTATTGATGCTTACCAGTCACAGCTCATCCAAATGACCCAGTT 600

Db 601 gctactgagtcagtcgtctcaggttcaaaagtgaatttcaaaagcatabcaggaaggg 660
 Qy 601 GTTCTGCTGCTGCTTGAATGTTTACCGGCTTCAATGGGTTAAACGAACTTCAGTTCAAACTTC 660

Qy 601 GCTACTGGAGTCATGGCTCTTCAAGTGAATTTCAAAGGCGCATACGAGAAAGGG 660

Db 661 attcaacaatacaaatattgggaacacacatatgaggaattccatgaattgattgctcaa 720
 Qy 661 ATTCACAAAATCAAAGTATTGGGAACCAACATATGAGGATTCATGAATCTGATTTGCTCAA 720

Db 721 gttccacttggctgcttattgtttatcgaggaatgtacaagaatggtgacactatacct 780
 Qy 721 GTTCCACTTGGTCTGCTTATGCTTTATCGCAGGATGTACAAGAATGGTGACACTATACCT 780

Db 781 aagatgaatccctggattatggtgcaaatttgtctcacatgctgtggttccagtagctct 840
 Qy 781 AAGGATGAATCCCTGGATTATGGTCAAAATTTGCTACATGCTTGTGTTTCAAGTAGCTCT 840

Db 841 gaaatcatgaactcttatgaggtctctatgaacaatacacagatgacatgaaggtggt 900
 Qy 841 GAAATCATGAACCTTCTTAAGAGCTCTATGTAACAATACACAGTGATCATGAAGGTGGI 900

Db 901 aatgtcagtgctcacacacggtcacttggctgtagtcttggctgacaccttaccctctcc 960
 Qy 901 AATGTCAAGTGTACACACCGTCACTTGGTGTGCTAGTGTGCTGTGATCCTTACCTCTCC 960

Db 961 tttgctgctgcttgaatggttttagccgaccacttcaagtttagccaatcagaagatt 1020
 Qy 961 TTTGCTGCTGCTTGAATGGTTTAGCCGACCACTTCAATGCTTTAGCCAAATCAGGAAAT 1020

Db 1021 ttgctatgataaaaatctgttgaagaatgtggggagaaacatttccaaagacagttg 1080
 Qy 1021 TTGCTATGATGAATATCTGTTGTAGAAGATGTGGGGAGAACATTTCCAAAGACAGTTG 1080

Db 1081 aaagactatgtttggaaaaaacattgacagtggaaggttgcctctggtttggacatgga 1140
 Qy 1081 AAAGACTATGTTTGGAAAAACATTGAACAGTGGCAAGGTGTCCCTGTTGTTGGACATGGA 1140

Db 1141 gttctcgaaacactgtaccagaatatacatgccagagaggttcgctataaagcatttg 1200
 Qy 1141 GTTCTCGGAAAGACTGTACCAAGATATACATGCCAGAGAGAGTTCGCTATGAAGCATTTG 1200

Db 1201 cctgaagatccactgtttcaactggtttcaaaactctcagagtttctcctcctgtctta 1260
 Qy 1201 CCTGAAGATCCACTGTTTCAACTGGTTCCTCAAACTCTACGAGTTTCTCCTCTCTTA 1260

Db 1261 cagaacttggcaaaagttaaaaccttggccaaatgttgatgccacagtggtggtgtgttg 1320
 Qy 1261 CAGAACTTGGCAAAAGTTAAACCTTGGCCAAATGTTGATGCCACAGTGTGTGTGTGTG 1320

Db 1321 aactattatggtttaaactgaagaacagatattatacgttccctcttggcggtatcaagact 1380
 Qy 1321 AACTATTATGGTTTAACTGAAGCAAGATATTATACGGTCTCTCTTTGGCTATCAAGAGCT 1380

Db 1381 cttggcatttgcctcagctaaattgggaccgagctcttgattgcccctagagaggcca 1440
 Qy 1381 CTTGGCATTTGCTCTCAGTAAATTTGGGACCGAGTCTCTTGGATTCTCGCTAGAGAGGCA 1440

Db 1441 aaagtgatcacaaatggaggttgagaccagtgcaagaagaatgaattgtttgaaat 1500
 Qy 1441 AAGAGTGTCACAATGGAGTGGCTTGAGAACCACTGAGCAAGCAATGAATTTGTTTGAAT 1500

Db 1501 ctgcgagcatataaacacaaatgataaattctctatgaataattgttgacaaagcactct 1560
 Qy 1501 CTGCGCAGCATAAACACAAATGTATAATCTCTATGAATAATTGCTTGACAAAGCAGTCT 1560

Db 1561 ttcttgggggacaagatagtgccgaccttcaatgggttaacgaacctcagttcaaatctc 1620
 Qy 1561 TTCTTGGGGGACAAGATAGTGGGCTTCAATGGGTTAAACGAACTTCAGTTCAAACTTC 1620

Db 1621 actgaattgtggaattgtattggtttctcagacacttctcgaattttgaacttagctct 1680
 Qy 1621 ACTGAATTTGTGGAATGTATGGTTTCTCGAGACTTGTCTGAAATTTTGAACCTTAGTCT 1680

Db 1681 agtggattcatctttcttctcctccgaattcctcacacactgatccagcatgtataaata 1740
 Qy 1681 AGTGGATTCTATTTCTTCTTCTTCCGAAATTCCTCACACGCTGATCCAGCATGTAAATA 1740

```

Db 1741 ataggtcaatgctattatgcggtctcttctggttgccattagactgtgtaacttctctt 1800
QY 1741 ATAGGTCAATGCTATTATATCGGGTCTCTTGGTTGCCATTAGACTTGTGAATGACTTCTTT 1800

Db 1801 gctgaaagttagtaatcggtgattcagcgaataaactgaattgttagttctttaa 1860
QY 1801 GCTGAAAGTTAGTAATCGGCTGATTTCAGCAATAAATGCAATGTTAGTTCTTTAAA 1860

Db 1861 ttgtctaattcttatttgatgatattatgaa 1891
QY 1861 TTGTCTAATCTTATTGTGATGATATTATGAA 1891

RESULT 3
ID T04201 standard: cDNA to mRNA: 1747 BP.
AC T04201;
DT 25-JAN-1996 (first entry)
DE Tobacco citrate synthase cDNA.
KW Citrate synthase; flower formation; ss.
OS Nicotiana tabacum
FH Key Location/Qualifiers
FT cds 70..1476
FT /*tag= a
PN W09524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE) HNECHST-SCHPEING AGREVN GMBH
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR P-PSDB: R82840.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Claim 7; Page 60-63; 87pp: English.
CC To identify a cDNA from tobacco which codes for citrate
CC synthase, a cDNA bank of leaf tissue from tobacco was prep'd.
CC Plaques of this cDNA bank were screened using a radioactive DNA
CC probe which comprises Solanum tuberosum citrate synthase cDNA
CC (T04199). One of the clones was sequenced. The nt. sequence is
CC given in T04201.
SQ Sequence 1747 BP; 490 A; 335 C; 400 G; 522 T;

Query Match 65.5%; Score 1238; DB 16; Length 1747;
Best Local Similarity 92.8%; Pred. No. 0.00e+00;
Matches 1439; Conservative 0; Mismatches 93; Indels 18; Gaps 14;

Db 38 ggtaaaagttaatttgtt-gattttgagagccatggtgtctctacgcggtttctctg 96
QY 40 GGTAAAAGTTAATTTTGTGATTTTCGCGAGCAATGGTGTCTACCGTAGCGTTTCGTTG 99

Db 97 ctgtcaaaactcgctctcgagcggttccacagacaaattctagcaactctgtcggtgg 156
QY 100 CTGTCAAAAGTCGCTCTCGAGCGGTCAACAGCAAAATGTTAGCAATTTCTGTCGCGTGG 159

Db 157 cttaagtcacaaactctctgtgtctgtatctgctgtctgagctg---caagaattgatt 213
QY 160 CTTCAGTCTCAAACTCTCTCGGCTCTTGTATCTGGGCTCTGAGCTGGTCAAGAATTTGAT 219

Db 214 ccagaacacagagatcgctaaagaagctcaagtcagagcatgaaaggttcaattgga 273
QY 220 CCTGAACAACAGGA:CGGCTGAAAAAGATCAAGTCAGAT-ATG-AAAGGTTCAATTGGGA 277

Db 274 ascatcaagtgataggtttgttggatgagagaatgacagagattactgtgggaa 333
QY 278 A-CATCACAGTTGATATGGTCTTCTGGTGGAAATGAGAGGAATGACAGGATTACTGTGG-AA 335

Db 334 acctcattac-ttgaccccgatgaagggaatcgcttttcggggcttgcctctctatgaatg 392
QY 336 ACCTCATTACCTTGACCCCTGATGAGGGAATTCGCTTCCGGGGGTTGTCTATACCTGAATG 395

```

```

Db 393 ccaaaaagattataactcgcagcaaaagcctggggagagaccccttcctgaaggtcttctctg 452
QY 396 CCAAAAAGTATTAGCTCGAGAAAAGCCTGGGGGTGAGCCCTTCGCTGAAGGCTTCTCTG 455

Db 453 gctctctttaacagaaaggtgcccatacaaaagacgaagtgttca-ttctctcaagaat 511
QY 456 GCTTCTTTTAAACAGAAAGGTGCCATCAAAAGACGAAGTGAATTCAAATTCATTTGCTCAGGAAT 515

Db 512 tgcgaagtcgtgtctac-tgtcccgcatcg-tatacaaaactattgatcccttaccagt 569
QY 516 TGCAGAGTGGGCGATCATATCCCTGATCATCATGTATACAACTATTGATGCTTTACCAAGT 575

Db 570 cacagctcatccaatgactcaagtttgctactgagtgatgctccttcagcttcaaaagtaa 629
QY 576 CACAGCTCATCCAATGACCCAGTTTGTACTGGAGTCAATCCCTGGATTATGGTGCAAAATTTTC 635

Db 630 attcaaaagcctatgagaagagattcacaatcaaaattatgggaacccgacatatga 689
QY 636 ATTCAAAAGGCTACGGAAGAGGATTTCANAATCAAAATATTTGGAAACCAACATATGA 695

Db 690 ggattccatgagtttgattgctcaagttccactgtgttgctgcttattgtttatccaggaat 749
QY 696 GGATTCCATGAATCTGATTGCTCAAGTTCACATGTTGCTGCTTATGTTTATCGCAGGAT 755

Db 750 gtacaagaacggcaacacactataccaaaggaatgactcaactgattatgggtcaaaatttgc 809
QY 756 GTACAAGAATGGTGACACTATACCTAAGGATGAATCCCTGGATTATGGTGCAAAATTTTC 815

Db 810 tccatgctgtgttctcagtagctctgacatgcatgagctt---atgaagctctatgtcac 866
QY 816 TCACATGCTTGGTTTCAGTAGCTCTGAAATGTCATGAACITCTTATGAGGCTCTATGTAAC 875

Db 867 gatacacagtgatcatgaaggggacgctcagtgctcacacaggtcacttggttactga 926
QY 876 AATACACAGTATCATGAAGGTGGTAATGTACAGTGTCTCACACCGGTCACCTTGGTGTCTAG 935

Db 927 tgccttgcagacccttacctctcctcgtcgtgctgttgaatggtttagctgagaccact 986
QY 936 TGCCTTGTCTCATCTTACCTCTCCTTGTGCTGCTGCTTTGAATGGTTTACCGGACCACT 995

Db 987 tcatggtttagccaatcaggaagttttgctatgagatacaatctgttgtagagagtgag 1046
QY 996 TCATGTTTAGCCAATCAGGAAGTTTTCGTATGGATAAAACTCTGTTGTAGACAAATGTGG 1055

Db 1047 ggagaaacatttccaaagagcagttgaaagactacgcttgaaaaacattgaaaagtggcaa 1106
QY 1056 GGAGAAACATTTCCAAAGAGCAGTTGAAAGACTATGTTTGGAAAAACATTTGAACAGTGGCAA 1115

Db 1107 ggtgtccctggttttcggacatggatggcttctgcgaagactgacccaagatcacacatgcca 1166
QY 1116 GGTGTCCCTGTTTGGACATGCAGTTCGCGAAAGACTGTACCAAGATATACATGCCA 1175

Db 1167 gagagagtcgcttgaagcaatttgcctgaagatcccaactgttcaactggttgcaaaact 1226
QY 1176 GAGAGAGTCTGCTATGAAGCATTTTCCTGAAGATCCACTGTTTCAACTGTTTCAAAAAC 1235

Db 1227 ctacgaactgttctccaattcttacaagaacttggcaaaagttaaaccttggcgcaaatgt 1286
QY 1236 CTACGAAGTTTTCCTCTCTGTTCTTACAGAACTTGGCAAAAGTTTAAAACCTTGGCCAAATGT 1295

Db 1287 tcatgccacagtggtgtgttggtaactattatggtttaactgaagcaagatattatatac 1346
QY 1296 TGATGCCACAGTGGTGTGTTGTTGAACATTATGTTTAACTGAAGCAAGATATTATATAC 1355

Db 1347 ggtcctctttagttatcaagagctcttggcatttgcctcagctcaatttgggaacgaagc 1406
QY 1356 GGTCTCTTTGGGCTATCAAGAGCTCTTGGCATTTGCTCTCAGCTAATTTTGGACCGAGC 1415

Db 1407 tcttgattgcccactagagagccaaagagtgatcaaatggatggcttgagaaacattg 1466
QY 1416 TCTTGATTGCGCTAGAGAGGCCAAAGAGTGTACAAATGAGTGGCTTGGAGAACCAAGTG 1475

```



```

FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; Z can be a
FT sequence of 6,9 or 12 nucleotides (see
FT comments)"
FT
FT WO9418318-A.
PD 18-AUG-1994.
PD 01-FEB-1994; U00977.
PD 01-FEB-1994; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
PI WPI: 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PT
PS Disclosure; Page 3; 255pp; English.
CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC This generic formula can be represented as follows: X(TGC)(NNB)10-
CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
CC sites (X is not the same as Y) that are not specified further. This
CC sequence generates peptides that are cloverleaf in structure. Other
CC generic sequences are shown in Q70465-68. Other specific peptides
CC generated by these generic sequences are shown in R65150-54. TSARs are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The oligonucleotides are also designed so that the expressed peptide
CC contains 2 or 4 cysteine residues positioned in, or flanking, the
CC unpredicted or variant residues. These residues confer some degree of
CC conformational rigidity to the peptides. The TSARs or compsns, comprising
CC a TSAR binding domain can be used in vivo to deliver a chemically or
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC or enzyme, to the specific target or on the cell. They can also replace
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC and therefore circumvent the need for complex methods of hybridoma
CC formation or in vivo antibody production. The TSARs are easily
CC characterised and have designed activity allowing direct and rapid
CC detection in a screening process.
CC
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 1.6%; Score 30; DB 12; Length 114;
Best Local Similarity 5.4%; Pred. No. 7.75e-04;
Matches 6; Conservative 30; Mismatches 76; Indels 0; Gaps 0;

Db 1 tgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 60
Cp 321 TGTCAATCTCTCAATCCACCAAGACCATATCAACTGTGATGTCTCCCAATGACCTTT 262
Db 61 nbnbnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnt 112
Cp 261 CATATCTGACTTGATCTTTTCAGGCATCTGTTGTTTCAGGAATCAATCT 210

RESULT 14
ID V44650 standard; DNA; 91 BP.
AC V44650;
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.
PI Cossons NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
DR WPI: 98-362770/31.
PT Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1; Page 42; 54pp; English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorsconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned

```

```

DR WPI: 98-362770/31.
PT Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1; Page 42; 54pp; English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorsconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
SQ Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

Query Match 1.5%; Score 28; DB 46; Length 91;
Best Local Similarity 15.3%; Pred. No. 9.56e-03;
Matches 13; Conservative 46; Mismatches 25; Indels 1; Gaps 1;

Db 3 mtwaakrawrwwkddavwvgaakrwwkwwhrasacmdkwaaktwkggtwarywgrk 62
QY 1462 CTTGAGAACAGTCGACAGAAAGCATGATTTGTAATCTCGGAGCATAAACACAAT 1521
Db 63 mwvtkwawdatakwkwwkdakwnw 87
QY 1522 GTATA-AICTATGATGAATATGCT 1545

RESULT 15
ID V44650 standard; DNA; 91 BP.
AC V44650;
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.
PI Cossons NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
DR WPI: 98-362770/31.
PT Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1; Page 42; 54pp; English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorsconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned

```

CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.

SQ Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

Query Match 1.5%; Score 29; DB 46; Length 91;

Best Local Similarity 13.3%; Pred. No. 2.75e-03;

Matches 10; Conservative 42; Mismatches 23; Indels 0; Gaps 0;

Db 13 wkkdvwgkwrkwwhrassacmdkkaaktwkggtwrrywkgrkmwttkawsd 72

Cp 1552 TTGTCAAGCAATTATTCATAGAGATTATACATTGTGTTTATGCTCGGAGATTTCAAA 1493

Db 73 atakwkkdskmw 87

Cp 1492 CAATCAATGCTTTCT 1478

Search completed: Sat Oct 23 16:31:37 1999
Job time : 429 secs.

CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: pTZ9pt-Fls
SQ SEQUENCE 7218 BP: 1944 A: 1491 C: 1486 G: 1929 T: 368 OTHER.

Query Match 2.4%; Score 45; DB 2; Length 7218;
Best Local Similarity 1.8%; Pred. No. 4.04e-16;
Matches 6; Conservative 179; Mismatches 140; Indels 0; Gaps 0;

Db 1116 YY 1175
Cp 527 CCGGACTCTGCAATCTCTGAGACATTAATCACTTCTCTTTTGTGGACCTTTCTCT 468
Db 1176 YY 1235
Cp 467 GTTAAAGCAAGCCAGAGAAGACCTTCAGGCAAGGCTCACCCAGGCTTTGCTCAGGT 408
Db 1236 YY 1295
Cp 407 AATACCTTTTGGCATTCAGGTATAGAACACCCGGAAGCGAATTCCTCATCAGGCTCA 348
Db 1296 YY 1355
Cp 347 AGTAATAGAGTTTCACAGTAATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 288
Db 1356 YY 1415
Cp 287 ACTGTGATGTCCCAATTCAGTATCATATCTGATCTGATCTTTTTCAGGCGATCTGT 228
Db 1416 YY 1480
Cp 227 TGTTCAGGAATCAATCTGTGTACCA 203

RESULT 2
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC XXXXX
DT

Sequence 14, Application US/08232463
DE Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935.313
CC FILING DATE:
CC APPLICATION NUMBER: Ep 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.

CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)883-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: pTZ9pt-Fls
SQ SEQUENCE 7218 BP: 1944 A: 1491 C: 1486 G: 1929 T: 368 OTHER.

Query Match 2.2%; Score 42; DB 2; Length 7218;
Best Local Similarity 1.9%; Pred. No. 5.99e-14;
Matches 4; Conservative 121; Mismatches 83; Indels 0; Gaps 0;

Db 1231 YY 1290
QY 1 TTTTGGTTCATCAGCCTACITGAGATGTATCCACATGGTAAAGTTAAITTTTCA 60
Db 1291 YY 1350
QY 61 TTTTCGGCAGCAATGGTGTCTTCTACCGTAGCGTTCTGCTGCAAGCTCCGCTCTCA 120
Db 1351 YY 1410
QY 121 GCGGTCCACAGTCAATGTAGCAATCTGTGCGTGGCTTCACTCAACCTCAACCTCTTCC 180
Db 1411 YYGTAC 1438
QY 181 GGTCTTGATCTGCTGCTGAGCTGGTAC 208

RESULT 3
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC XXXXX
DT

Sequence 22, Application US/08388672A
DE Sequence 22, Application US/08388672A
CC Patent No. 5795961
CC GENERAL INFORMATION:
CC APPLICANT: Wallace, T. Paul
CC APPLICANT: Harris, William J.
CC APPLICANT: Carr, Frank J.
CC APPLICANT: Old, Lloyd J.
CC APPLICANT: Welt, Sydney
CC APPLICANT: Kitamura, Kunio
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Felfe and Lynch
CC STREET: 805 Third Avenue
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10022
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/388,672A
CC FILING DATE: 14-FEB-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, No. 5795961man D.
CC REGISTRATION NUMBER: 30,946

CC REFERENCE/DOCKET NUMBER: LUD 5409
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-688-9200
CC TELEFAX: 212-838-3884
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 1.8%; Score 34; DB 3; Length 965;
Best Local Similarity 17.3%; Pred. No. 2,31e-08;
Matches 14; Conservative 42; Mismatches 25; Indels 0; Gaps 0;

Db 851 TAVYCVGRSVDSDGDYWGTTVTVSSHUVKDMTSSASVSGRVTTCRSTTHNGN 910
QY 1400 TAAITTTGGACCGAGCTCTTGATTCGCGTAGAGAGCCAAAGAGTGTCACAATGGAGT 1459
Db 911 TYVYKGYKAKYRVNSRGSVR 931
QY 1460 GCGTTGAGACCAAGTCAAGA 1480

RESULT 4
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC xxxxxx
DT
DE Sequence 5, Application US/08238163
CC Sequence 5, Application US/08238163
CC Patent No. 5569830
CC GENERAL INFORMATION:
CC APPLICANT: BENNETT, Alan
CC APPLICANT: LABAVITCH, John M.
CC APPLICANT: POWELL, Ann
CC APPLICANT: STOTZ, Henrik
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL
CC TITLE OF INVENTION: POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend Khourie and Crew
CC STREET: Steuart Street Tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/238,163
CC FILING DATE: 03-MAY-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc_feature

CC LOCATION: 1..215
CC OTHER INFORMATION: /standard_name= "Deduced amino acid
CC OTHER INFORMATION: Sequence of FGIP from bean."
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 1.6%; Score 31; DB 1; Length 215;
Best Local Similarity 14.0%; Pred. No. 2.30e-06;
Matches 28; Conservative 75; Mismatches 96; Indels 1; Gaps 1;

Db 9 SSVVSRASCNDRAKKGNTTSSWTTDCCNRTWGVCDTDTTYRVNDSGHKYSANYNY 68
QY 967 GCTGCTTTGAATGGTTTAGCGGACCACTTCATGTTAGCCAAATCAGGAAGTTTGTCTA 1026
Db 69 GGNVGAATHYTHTVSGADSKTVTDNSYNSGTTDGNRSGADSYGSSKTAMT 128
QY 1027 TGGATAAATCTGTGTAGAGAATGTGGGAGAACATTTCCAAAGACAGTGAAGAC 1086
Db 129 SRNRTGTANNVDSRNMGDASVSGDKNTKKHAKNSADGKVGSKNNDRNNRYGTGTSN 188
QY 1087 TATGT-TTGGAAAACATTGAACAGTGGCAAGTTGTCCCTGGTTTGGACATGGAGTCT 1145
Db 189 VSNCGGKGRDVSYYANK 208
QY 1146 GCGAAAGACTGTACCAAGAT 1165

RESULT 5
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC xxxxxx
DT
DE Sequence 5, Application US/08238163
CC Sequence 5, Application US/08238163
CC Patent No. 5569830
CC GENERAL INFORMATION:
CC APPLICANT: BENNETT, Alan
CC APPLICANT: LABAVITCH, John M.
CC APPLICANT: POWELL, Ann
CC APPLICANT: STOTZ, Henrik
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL
CC TITLE OF INVENTION: POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend Khourie and Crew
CC STREET: Steuart Street Tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/238,163
CC FILING DATE: 03-MAY-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc_feature

CC LOCATION: 1, 215
CC OTHER INFORMATION: /standard_name= "Deduced amino acid
CC OTHER INFORMATION: sequence of PGIP from bean."
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 1.6%; Score 31; DB 1; Length 215.
Best Local Similarity 11.5%; Pred. No. 2,30e-06;
Matches 24; Conservative 80; Mismatches 104; Indels 0; Gaps 0;

Db 1 MTWVSSSVSRATSCNDKAKKDGNTTSSWTDCNRTWGVCDTDTIRVNNDSGHNK 60
Cp 291 ATCAACTGTGATGCCCAATGAACCTTTCATATGCTACCTGCTTTTTCAGCGGATC 232
Db 61 YSSANYNGNNVGAATHYHTHTNVSGADSKTVTDSYNASGTSSTSSNGTDCNRSADSY 120
Cp 231 CTGCTTTCAGGATCAATCTTGTACCAAGCTCAGAACCCAGATCAACACCGGAAGGT 172
Db 121 GSKTAMTSRNRGTANNVADSRNMGDASVGSNDKNTKKHAKNSADGKVGSKNNGDRNR 180
Cp 171 TTGACTTGAAGCCAGCCGACAGAAATGCTAACATTTGACTCTTGACCGCTCGAGAGCG 112
Db 181 YGTGTSNVSNCGGNKRDVSSYANK 208
Cp 111 GAGCTTTGACAGCAAGAAACGCTACGG 84

RESULT 6
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC xxxxxx
DT
DE Sequence 22, Application US/08388672A
DE Sequence 22, Application US/08388672A
CC Patent No. 5795961
CC GENERAL INFORMATION:
CC APPLICANT: Wallace, T. Paul
CC APPLICANT: Harris, William J.
CC APPLICANT: Carr, Frank J.
CC APPLICANT: Old, Lloyd J.
CC APPLICANT: Welt, Sydney
CC APPLICANT: Kitamura, Kunio
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B
CC TITLE OF INVENTION: Antibodies
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Felfe and Lynch
CC STREET: 805 Third Avenue
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10022
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/388,672A
CC FILING DATE: 14-FEB-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, No. 5795961man D.
CC REGISTRATION NUMBER: 30,946
CC REFERENCE/DOCKET NUMBER: LUD 5409
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-688-9200
CC TELEFAX: 212-838-3884
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: DNA (genomic)

SQ SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 1.5%; Score 28; DB 3; Length 965;
Best Local Similarity 21.3%; Pred. No. 1.95e-04;
Matches 19; Conservative 38; Mismatches 32; Indels 0; Gaps 0;

Db 836 ADSSNSRSSVTAADTAVYVCVRGRSVDSGDYWGTTVTVSSHUHKMTSSSSASVGD 895
Cp 1589 AAGGCCGACCTATCTTGTCCCCAAGAGAGTGCTTGTCAAGCAATTATTATAGATA 1530
Db 896 RVITCRSSTTHGNTYTYWYKGRKAYRVS 924
Cp 1529 GATTATACATGTGTTTATGCTCGCGAG 1501

RESULT 7
ID US-08-529-600D-3 STANDARD; DNA; UNC; 1302 BP.
AC xxxxxx
DT
DE Sequence 3, Application US/08529600D
DE Sequence 3, Application US/08529600D
CC Patent No. 5861285
CC GENERAL INFORMATION:
CC APPLICANT: Tadashi MATSUNAGA
CC TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND
CC TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FILLSBURY MADISON & SUTRO, L.L.P.
CC STREET: 1100 NEW YORK AVENUE, N.W.
CC CITY: WASHINGTON
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005-3918
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: JP 6-248700
CC FILING DATE: 16-SEP-1994
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1302 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 1302 BP; 143 A; 118 C; 259 G; 216 T; 566 OTHER.

Query Match 1.5%; Score 28; DB 3; Length 1302;
Best Local Similarity 32.2%; Pred. No. 1.95e-04;
Matches 49; Conservative 26; Mismatches 76; Indels 1; Gaps 1;

Db 704 TNYTNGAYYTNWSCNCGNCTAYGGNCGNTYTYTNGNGGNGINGTNYTNGNAAVWSNG 763
Cp 1646 TTCTCGAGACTTGTCTGTAATTTTGAACCTAGTCTAGTGAGTCAATTTTCTTCATCCG 1705
Db 754 CNCARMGNGAVATGYTNYTNAARMGNCNCARCNCNATHGNGWNGNTNYTNYTATGCTNT 823
Cp 1706 AATCCCTC-ACAGCTGATCCAGCATGTAAATAATTAAAGTCAATGCTAATTAATCGCT 1764
Db 824 TTYTYTNWSNATHGGNYTNYTNYTNGAYTTY 855
Cp 1765 TCTTGTGTGCAATAGACTTGTGAATGACTTC 1796

RESULT 8

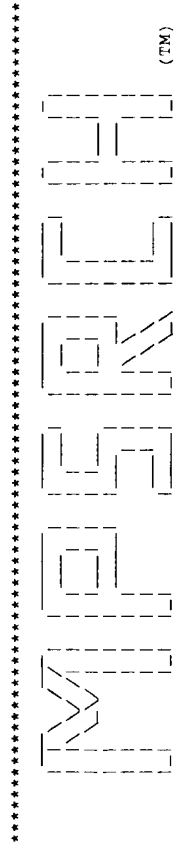

```
CC APPLICANT: Little, Melvyn
CC APPLICANT: Breitting, Frank B
CC APPLICANT: Seehaus, Thomas
CC APPLICANT: Dubel, Stefan
CC APPLICANT: Klewinghaus, Iris
CC TITLE OF INVENTION: Preparation and Use of Gene Banks of
CC TITLE OF INVENTION: Synthetic Human Antibodies ("Synthetic Human-Antibody
CC TITLE OF INVENTION: Libraries")
CC NUMBER OF SEQUENCES: 39
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/353,372A
CC FILING DATE: 02-DEC-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/247,393
CC FILING DATE: 23-MAY-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/654,207
CC FILING DATE: 30-JAN-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: DE P 40 02 897.6
CC FILING DATE: 01-FEB-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: DE P 40 03 880.7
CC FILING DATE: 09-FEB-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fortan, David S
CC REGISTRATION NUMBER: 33,694
CC FEEPENCODE/DOCKET NUMBER: 05552.1032-02000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 408-4000
CC TELEFAX: (202) 408-4400
CC INFORMATION FOR SEQ ID NO: 16:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 92 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 92 BP; 13 A; 10 C; 17 G; 17 T; 35 OTHER.

Query Match 1.3%; Score 24; DB 3; Length 92;
Best Local Similarity 17.8%; Pred.No. 5.36e-02;
Matches 8; Conservative 24; Mismatches 13; Indels 0; Gaps 0;

Db 23 BAKCNBNHNKNNRPTPPCFVTAHAPYMTATPYCSMMAMVBTCVD 67
QY 753 GATGTACAAAGTGGTGACTATACCCTAAGGATGAATCCTCGGA 797
:::|||||:::|||||:::|||||:::|||||:::

RESULT 14
ID US-07-865-662F-13 STANDARD; DNA; UNC; 105 BP.
AC xxxxxx
DT
DE Sequence 13, Application US/07865662F
CC Sequence 13, Application US/07865662F
CC Patent No. 5451670
CC GENERAL INFORMATION:
```

CC	APPLICANT:	Marcia M. Miller
CC	TITLE OF INVENTION:	Restriction Fragment Length
CC	TITLE OF INVENTION:	Polymorphism Test For Haplotyping Domesticated Fowl
CC	NUMBER OF SEQUENCES:	14
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	City of Hope
CC	STREET:	1500 East Duarte Road
CC	CITY:	Duarte
CC	STATE:	California
CC	COUNTRY:	United States of America
CC	ZIP:	91010-0269
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	3M Double Density 5 1/4" diskette
CC	OPERATING SYSTEM:	MS DOS Version 3.20
CC	SOFTWARE:	Microsoft
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/07/865,662F
CC	FILING DATE:	07 April, 1992
CC	CLASSIFICATION:	435
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	07/688,326
CC	FILING DATE:	22 April 1991
CC	APPLICATION NUMBER:	07/588,922
CC	FILING DATE:	27 September 1990
CC	APPLICATION NUMBER:	07/210,405
CC	FILING DATE:	23 June 1988
CC	APPLICATION NUMBER:	US 07/130,529
CC	FILING DATE:	9 December 1987
CC	APPLICATION NUMBER:	US 07/068,176
CC	FILING DATE:	30 June 1987
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Irons, Edward S.
CC	REGISTRATION NUMBER:	16,541
CC	REFERENCE/DOCKET NUMBER:	No. 5451670e
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(202) 783-6040
CC	TELEFAX:	(202) 783-6031
CC	TELEX:	No. 5451670e
CC	INFORMATION FOR SEQ ID NO:	13:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	105
CC	TYPE:	Nucleic Acid
CC	STRANDEDNESS:	Double
CC	TOPOLOGY:	Linear
CC	MOLECULE TYPE:	DNA
CC	ORIGINAL SOURCE:	Synthetically Prepared
CC	IMMEDIATE SOURCE:	Synthetically Prepared
CC	SEQUENCE 105 BP; 15 A; 0 C; 8 G; 1 T; 81 OTHER.	
QY	Query Match	1.3%; Score 24; DB 1; Length 105;
QY	Best Local Similarity	13.0%; Fred.No. 5.36e-02;
QY	Matches	12; Conservative 27; Mismatches 53; Indels 0; Gaps 0;
Db	12	SNNKSNVAVKNGNNAKNKNSAKNSAANVNNGNNAKNKNSNNMCTPKNPNAAKN 71
QY	416	CAAGCCTGGGGGTGAGCCCTTCGCCTSAAGTGTTCTCGGTCTTCTTTTAACAGGAAGG 475
Db	72	NSGVADNKKNASNNYDNGSGVADNKNAKNKY 103
QY	476	TGCCATCAAAGAAGCAAGTGAATTCAATTGTC 507
RESULT	15	
ID	US-08-137-175A-7	STANDARD: DNA; UNC: 1958 BP.
AC	xxxxxx	
DE	Sequence 7,	Application US/08137175A
CC	Sequence 7,	Application US/08137175A
CC	Patent No. 5777095	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	BARBOUR, Alan G.
CC	APPLICANT:	BERGSTROEM, Sven



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 16:36:55 1999; MasPar time 7.33 Seconds
Tabular output not generated. 752.832 Million cell updates/sec

Title: >US-08-702-718-2
Description: (1-471) from US08702718.pep
Perfect Score: 3466
Sequence: 1 MVFYRSVLSKLSRAVQ.....PLRPKSVTMEWLENOCKKA 471

Scoring table: PAM 150
Gap 11

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 33.654; Variance 162.896; scale 0.207

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	113	3.3	220	1	US-08-063- Sequence 10, Applicati	2.42e+00
2	113	3.3	220	3	PCT-US93-0 Sequence 10, Applicati	2.42e+00
3	115	3.3	3031	1	US-07-689- Sequence 2, Applicati	1.77e+00
4	100	2.9	268	4	5320958-6 Patent No. 5320958.	1.79e+01
5	97	2.8	355	1	US-08-450- Sequence 5, Applicatio	2.80e+01
6	97	2.8	355	3	PCT-US95-0 Sequence 5, Applicatio	2.80e+01
7	97	2.8	355	1	US-08-012- Sequence 2, Applicati	2.80e+01
8	92	2.7	98	2	US-08-345- Sequence 29, Applicati	5.85e+01
9	92	2.7	108	2	US-08-245- Sequence 10, Applicati	5.85e+01
10	93	2.7	225	2	US-08-637- Sequence 127, Applicat	5.05e+01
11	94	2.7	587	1	US-08-844- Sequence 2, Applicatio	4.36e+01
12	92	2.7	1548	2	US-08-463- Sequence 7, Applicatio	5.85e+01
13	92	2.7	1548	2	US-08-460- Sequence 7, Applicatio	5.85e+01
14	89	2.6	485	4	5320958-2 Patent No. 5320958.	9.04e+01
15	90	2.6	763	4	US-08-742- Sequence 4, Applicatio	7.82e+01
16	91	2.6	771	2	US-08-742- Sequence 148, Applicat	6.77e+01
17	91	2.6	1312	2	US-08-592- Sequence 36, Applicati	6.77e+01
18	89	2.6	2265	2	US-08-149- Sequence 2, Applicatio	9.04e+01
19	86	2.5	170	2	US-08-818- Sequence 2, Applicatio	1.39e+02
20	85	2.5	335	2	US-08-761- Sequence 2, Applicatio	1.60e+02
21	87	2.5	340	2	US-08-933- Sequence 26, Applicati	1.20e+02
22	85	2.5	509	2	US-08-737- Sequence 27, Applicati	1.60e+02
23	85	2.5	550	1	US-08-279- Sequence 16, Applicati	1.60e+02

ALIGNMENTS

RESULT	1	US-08-063-552-10	STANDARD;	PRT;	220 AA.
ID	US-08-063-552-10	Sequence 10, Application US/08063552			
XX	xxxxxx	Sequence 10, Application US/08063552			
AC		Patent No. 5688936			
XX		GENERAL INFORMATION:			
CC		APPLICANT: Edwards, Robert H			
CC		TITLE OF INVENTION: Vesicle Membrane Transport Proteins			
CC		NUMBER OF SEQUENCES: 17			
CC		CORRESPONDENCE ADDRESS:			
CC		ADDRESSEE: Sheldon & Mak			
CC		STREET: 225 South Lake Avenue, Ninth Floor			
CC		CITY: Pasadena			
CC		STATE: California			
CC		COUNTRY: USA			
CC		ZIP: 91101			
CC		COMPUTER READABLE FORM:			
CC		MEDIUM TYPE: Floppy disk			
CC		COMPUTER: IBM PC compatible			
CC		OPERATING SYSTEM: PC-DOS/MS-DOS			
CC		SOFTWARE: PatentIn Release #1.0, Version #1.25			
CC		CURRENT APPLICATION DATA:			
CC		APPLICATION NUMBER: US/08/063,552			
CC		FILING DATE: 19930514			
CC		CLASSIFICATION: 530			
CC		ATTORNEY/AGENT INFORMATION:			
CC		NAME: Farber, Michael B			
CC		REGISTRATION NUMBER: 32,612			
CC		REFERENCE/DOCKET NUMBER: 9067-1			
CC		TELECOMMUNICATION INFORMATION:			
CC		TELEPHONE: (818) 796-4000			
CC		TELEFAX: (818) 795-6321			
CC		INFORMATION FOR SEQ ID NO: 10:			
CC		SEQUENCE CHARACTERISTICS:			
CC		LENGTH: 220 amino acids			
CC		TYPE: AMINO ACID			
CC		TOPOLOGY: linear			
CC		MOLECULE TYPE: peptide			
CC		HYPOTHETICAL: NO			
CC		FRAGMENT TYPE: internal			
CC		ORIGINAL SOURCE:			

CC	ORGANISM:	Rattus rattus
SQ	SEQUENCE	220 AA; 23695 MW; 270630 CN;
	Query Match	3.3%; Score 113; DB 1; Length 220;
	Best Local Similarity	20.2%; Pred No. 2.42e+00;
Matches	22; Conservative	41; Mismatches 41; Indels 5; Gaps 5;
Dbb	24 DNMLLTVVVPVPTFLYATEFKDSNSLHRGSPVSQSENVRIGILFASKALMOLLVNPF	83
QY	209 DSMNLIAQVPLVAAYVRYRMKNGDT-IPKDESLDYGANFAHM-LGFSSSEMHELLMRLY	266
Dbb	84 VGPLNIRGIYPHVFEGFMIMF-LSTLMFAFSGTVALLFVARTLOGIS	131
QY	267 VTIHSDHEGGNVSAHTCHLVASALSOPYLSFAAALNGL-AG-PLHGLAN	313
RESULT	2	
ID	PCT-US93-05704-10	STANDARD; PRT; 220 AA.
XX	xxxxxx	
AC		
AT		
DT		
DE	Sequence 10, Application PC/TUS9305704	
XX		
CC	sequence 10, Application PC/TUS9305704	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Edwards, Robert H	
CC	TITLE OF INVENTION: Vesicle Membrane Transport Proteins	
CC	NUMBER OF SEQUENCES: 17	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Sheldon & Mak	
CC	STREET: 225 South Lake Avenue, Ninth Floor	
CC	CITY: Pasadena	
CC	STATE: California	
CC	COUNTRY: USA	
CC	ZIP: 91001	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: Patent In Release #1 0, Version #1.25	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: PCT/US93/05704	
CC	FILING DATE: 19930611	
CC	CLASSIFICATION:	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Farber, Michael B	
CC	REGISTRATION NUMBER: 32.612	
CC	REFERENCE/DOCKET NUMBER: 9067-1PCT	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (818) 796-4000	
CC	TELEFAX: (818) 795-6321	
CC	INFORMATION FOR SEQ ID NO: 10:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 220 amino acids	
CC	TYPE: AMINO ACID	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: Peptide	
CC	HYPOTHETICAL: NO	
CC	FRAGMENT TYPE: internal	
CC	ORIGINAL SOURCE:	
CC	ORGANISM: Rattus rattus	
SQ	SEQUENCE	220 AA; 23695 MW; 270630 CN;
	Query Match	3.3%; Score 113; DB 3; Length 220;
	Best Local Similarity	20.2%; Pred No. 2.42e+00;
Matches	22; Conservative	41; Mismatches 41; Indels 5; Gaps 5;
Dbb	24 DNMLLTVVVPVPTFLYATEFKDSNSLHRGSPVSQSENVRIGILFASKALMOLLVNPF	83
QY	209 DSMNLIAQVPLVAAYVRYRMKNGDT-IPKDESLDYGANFAHM-LGFSSSEMHELLMRLY	266

Db 477 LYRVITIIILFSPKFNVTGKGVLEBEEDLGTATYNI--EAGIMTLGLLIGLFTF 535
QY 214 IAOVPLVA--AYVRYRMKNGDT--IPKQESUDYGANFAHMLGFSSEMHELMRLI-VTI 269
Db 536 HFNQLAG 542
QY 270 HSDHEGG 276
RESULT 4
ID 5320958-6 STANDARD; PRT; 290 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Patent No. 5320958.
XX
CC APPLICANT: INOUE, SUNIKO; HSU, MEI-YIN; EAGLE, SUSAN;
CC INOUE, MASAYORI
CC TITLE OF INVENTION: ISOLATED BACTERIAL REVERSE TRANSCRIPTASE
CC NUMBER OF SEQUENCES: 24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/315,316
CC FILING DATE: 24-FEB-1989
CC SEQ ID NO: 6
CC LENGTH: 268
CC SEQUENCE 290 AA; 32072 MW; 450220 CN;
Query Match 2.9%; Score 100; DB 4; Length 268;
Best Local Similarity 29.2%; Pred. No. 1.79e+01;
Matches 35; Conservative 25; Mismatches 51; Indels 9; Gaps 7;
Db 35 KAQRWVLSNVYERLPVGHGAHGFVAGRSILTNALAHQAGVAVVVKVLDLKDFPSPVWRRV 94
QY 16 RAQQQNSVNSVNWLVQVQTS--GL-DLSELVQELPEQQDRKKIK-SDMKGSIGNITV 72
Db 95 KGLLRKGLREGTSTLLSLLSTEAPRAVQFRGKLLHVAKGPRALPQGAPTTSPGITNAL 154
QY 73 DMVL--GGMR-GMTGLLWPHYLDPDGIRFRG--LSIPECCOKVLPAAKPGGEP-LPEGL 126
RESULT 5
ID US-08-450-393A-5 STANDARD; PRT; 355 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application US/08450393A
XX
CC Sequence 5, Application US/08450393A
CC Patent No. 5707815
CC GENERAL INFORMATION:
CC APPLICANT: Charo, Israel
CC APPLICANT: Coughlin, Shaun
CC TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMAATTRACTANT
CC TITLE OF INVENTION: PROTEIN RECEPTORS
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
CC STREET: 5 Palo Alto Square
CC CITY: Palo Alto
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94306-2155
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/450,393A
CC FILING DATE: May 25, 1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Cseri, Luann
CC REGISTRATION NUMBER: 31,822
CC REFERENCE/DOCKET NUMBER: UCAL-237/02US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-843-5165
CC TELEFAX: 415-8857-0663
CC TELEX: 380816COOLEYPA
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 355 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC SEQUENCE 355 AA; 41172 MW; 710742 CN;
Query Match 2.8%; Score 97; DB 1; Length 355;
Best Local Similarity 31.9%; Pred. No. 2.80e+01;
Matches 23; Conservative 14; Mismatches 29; Indels 5; Gaps 6;
Db 65 LKNMTSYLLNLAIISDLLFLFTLPFWIDY-KLKDDWVFGDAMCKILSGFYTYGLSEIFF 123
QY 373 MKHLPEPDLFOL-VSKLYEVFELL-FLQNLAKLP-WPNVDAHSGVLIN-YY-GLTEARYY 427
Db 124 IILLTDIYLA 135
QY 428 TVLFGVSRLGI 439
RESULT 6
ID PCT-US95-00476-5 STANDARD; PRT; 355 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application PC/TUS9500476
XX
CC Sequence 5, Application PC/TUS9500476
CC GENERAL INFORMATION:
CC APPLICANT: The Regents of the University of California
CC TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMAATTRACTANT
CC TITLE OF INVENTION: PROTEIN RECEPTORS
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Robbins, Berliner & Carson
CC STREET: 201 N. Figueroa Street, 5th Floor
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: USA
CC ZIP: 90012-2628
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/00476
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Berliner, Robert
CC REGISTRATION NUMBER: 20,121
CC REFERENCE/DOCKET NUMBER: 5555-291
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 310-977-1001
CC TELEFAX: 310-977-1003
CC TELEX:

100


```

XX Sequence 7, Application US/08460907B
DE Sequence 7, Application US/08460907B
XX Patent No. 5891724
CC GENERAL INFORMATION:
CC APPLICANT: Delevy, Roger G.
CC APPLICANT: Cole, Susan P.C.
CC TITLE OF INVENTION: METHODS FOR CONFERRING MULTIDRUG
CC TITLE OF INVENTION: RESISTANCE ON A CELL
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
CC STREET: Kingston
CC CITY: Kingston
CC STATE: Ontario
CC COUNTRY: CANADA
CC ZIP: K7L 3N6
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII text
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/460,907B
CC FILING DATE: 05-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/966,923
CC FILING DATE: 27-OCT-1992
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/029,340
CC FILING DATE: 8-MAR-1993
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/141,893
CC FILING DATE: 26-OCT-1993
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/407,207
CC FILING DATE: 20-MAR-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Steeg, Carol Miernicki
CC REGISTRATION NUMBER: 39,539
CC REFERENCE/DOCKET NUMBER: Q1551
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (613) 545-2342
CC TELEFAX: (613) 545-6853
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1548 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 1548 AA; 172282 MW; 12743552 CN;

Query Match 2.7%; Score 92; DB 2; Length 1548;
Best Local Similarity 28.2%; Pred. No. 5.85e+01;
Matches 22; Conservative 18; Mismatches 34; Indels 4; Gaps 4;

Db 370 VGLRVFGMAVFLVTLPL-QAVISKHVQDYSERMASVVDLRI-KRTNEL-LSGVRIVKFM 426
Qy 27 VFWLQVOTSSGLDRLSELVDELPIEQ-QDPLKIKSDMKSGISNTIVDVLGSGMRGTGL 85
Db 427 GWPEVFLARIODARSREL 444
Qy 86 LWKPHYLDPOEGIRFGL 103

RESULT 14

```

```

ID 5320958-2 STANDARD; PRT; 525 AA.
XX
AC xxxxxx
XX
XX 01-JAN-1900
XX Patent No. 5320958.
XX
CC Patent No. 5320958
CC APPLICANT: INOUE, SUNIKO; HSU, MEI-YIN; EAGLE, SUSAN;
CC INOUE, MASAYORI
CC TITLE OF INVENTION: ISOLATED RACTEPIAL REVERSE TRANSCRIPTASE
CC NUMBER OF SEQUENCES: 24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/315,316
CC FILING DATE: 24-FEB-1989
CC SEQ ID NO: 2:
CC LENGTH: 485
CC SEQUENCE 525 AA; 57562 MW; 1459426 CN;

Query Match 2.6%; Score 89; DB 4; Length 485;
Best Local Similarity 26.9%; Pred. No. 9.04e+01;
Matches 32; Conservative 27; Mismatches 52; Indels 8; Gaps 6;

Db 204 KAAQRWLSNVVERLPHVGAHGFVAGRSILTNALAHQAGADVVKDKDFPFSVTWRRV 263
Qy 16 RAVQOSNVSNVSRVWLOVOTSS-GL-DLRSELVQELIPEODRLKIK-SDMKSGISNTIV 72
Db 264 KGLLRKGGLEGTSTLLSLSTEAPRONVQPRGKLLHVAKGPRALPOGAPTSPOITNAL 322
Qy 73 DMVL--GGMR-GMTGLLWKPHYLDPOEGIRFG--LSIPEQKYLPAKPGGEPLPEGL 126

RESULT 15
ID US-08-742-753-4 STANDARD; PRT; 763 AA.
XX
AC xxxxxx
XX
XX
XX
XX Sequence 4, Application US/08742753
XX Sequence 4, Application US/08742753
XX Patent No. 5861278
XX GENERAL INFORMATION:
XX APPLICANT: WONG, Gordon G.
XX APPLICANT: YAO, Kwok-Ming.
XX TITLE OF INVENTION: HNF3-delta Compositions
XX NUMBER OF SEQUENCES: 4
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Genetics Institute, Inc.
XX STREET: 87 CambridgePark Drive
XX CITY: Cambridge
XX STATE: Massachusetts
XX COUNTRY: USA
XX ZIP: 02140
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: Patentin Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/742,753
XX FILING DATE:
XX CLASSIFICATION: 514
XX ATTORNEY/AGENT INFORMATION:
XX NAME: LAZAR, Steven R.
XX REGISTRATION NUMBER: 32,618
XX REFERENCE/DOCKET NUMBER: 5277
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: (617) 498-8260
XX TELEFAX: (617) 876-5851
XX INFORMATION FOR SEQ ID NO: 4:

```

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 763 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 70..2358
SQ SEQUENCE 763 AA; 84252 MW; 3182536 CN;

Query Match 2.6%; Score 90; DB 2; Length 763;
Best Local Similarity 27.1%; Pred. No. 7.82e+01;
Matches 19; Conservative 21; Mismatches 26; Indels 4; Gaps 4;

Db 334 LPEHLESOQKRPNPDELRRNM-TIKTEPLGARRKMKPDLPRVSSYLVPIDPFVNCSLVLQ 392
Qy 49 IPEQ-QDRLKKIKSDMKSGNITVDWVGGMRGTMGLL-WKPHYLDPDDEGIFRGLSIP 106
Db 393 PSVKVPLPLA 402
Qy 107 ECQKV-LPAA 115

Search completed: Fri Oct 22 16:38:18 1999
Job time : 83 secs.

WQISREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 16:20:42 1999; MasPar time 18.71 Seconds
Tabular output not generated. 535.432 Million cell updates/sec

Title: >US-08-702-718-2

Description: (1-471) from US08702718.pap

Perfect Score: 3466

Sequence: 1 MVFYRSVLLSLKRSRAVQ.....PLRPKSVTMEWLENQCKKA 471

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.074; Variance 166.895; scale 0.216

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3466	100.0	471	15	Potato citrate synth	0.00e+00
2	3466	100.0	471	14	Potato citrate synth	0.00e+00
3	3098	89.4	469	14	Tobacco citrate synth	1.27e-270
4	2637	76.1	437	14	Sugar beet citrate sy	8.02e-228
5	310	8.9	436	3	Citrate synthase enco	1.16e-15
6	130	3.8	754	35	Acetobacter xylinum b	2.91e-01
7	113	3.3	220	9	Fragment of chromaffi	4.63e+00
8	115	3.3	754	8	Cellulose synthase op	3.37e+00
9	110	3.2	3084	35	Mouse laminin A chain	7.44e+00
10	105	3.0	355	24	Rat CC chemokine rece	1.63e-01
11	100	2.9	329	29	H. pylori ORF 06ep302	3.50e-01
12	100	2.9	329	29	H. pylori ORF hp3e110	3.50e-01
13	101	2.9	456	34	Methyl tetrahydropter	3.01e-01
14	100	2.9	4473	19	Virulence gene cluste	3.50e+01
15	97	2.8	355	25	Human MIP-1 alpha/RAN	5.51e+01
16	97	2.8	355	11	C-C chemokine recepto	5.51e+01

17	97	2.8	355	24	W25751	Human MIP-1alpha/RANT	5.51e+01
18	96	2.8	365	13	R69518	Prostaglandin-EP3-9 r	6.40e+01
19	96	2.8	388	13	R69517	Prostaglandin-EP3-21	6.40e+01
20	96	2.8	390	13	R69516	Prostaglandin-EP3-alp	6.40e+01
21	96	2.8	393	32	W57411	Human prostaglandin E	6.40e+01
22	96	2.8	402	32	W57410	Human EP3-v receptor	6.40e+01
23	92	2.7	108	14	R70156	Streptococcus pneumon	1.16e+02
24	94	2.7	587	27	W31916	Streptococcus pneumon	8.63e+01
25	89	2.6	94	32	W54442	Mouse novel secreted	1.80e+02
26	90	2.6	107	35	W70233	Leishmania antigen Lc	1.55e+02
27	89	2.6	336	4	R20812	T-lymphocyte TLISA An	1.80e+02
28	89	2.6	336	18	W00860	Human T-lymphocyte TL	1.80e+02
29	89	2.6	336	23	W18051	DNAX accessory molecu	1.80e+02
30	91	2.6	419	24	W27464	B. cepacia insertion	1.34e+02
31	89	2.6	485	2	R11919	Reverse transcriptase	1.80e+02
32	90	2.6	757	33	W60486	Mouse TRIDENT transcr	1.55e+02
33	90	2.6	763	32	W60591	Human hepatocyte nucl	1.34e+02
34	91	2.6	771	32	W60590	Rat hepatocyte nuclea	1.34e+02
35	91	2.6	1312	35	W71295	Human homologue of ye	1.34e+02
36	91	2.6	1312	36	W22775	Human RAD50	1.34e+02
37	89	2.6	3011	4	R22154	NANBV Hutch c59 isola	1.80e+02
38	87	2.5	143	35	W38734	Streptococcus pneumon	2.40e+02
39	88	2.5	393	32	W60133	M. vaccae potd homolo	2.07e+02
40	87	2.5	399	32	W37707	Mouse GRP protein seq	2.40e+02
41	88	2.5	479	23	W22341	P. membranaefaciens N	2.07e+02
42	88	2.5	968	28	W41310	CF-5 pathogen resista	2.07e+02
43	88	2.5	968	28	W41309	CF-5 pathogen resista	2.07e+02
44	88	2.5	1016	28	W41312	CF-5 pathogen resista	2.07e+02
45	88	2.5	1172	31	W37870	Human protein compris	2.07e+02

ALIGNMENTS

RESULT 1
ID R86383 standard; Protein; 471 AA.

AC R86383;
DT 24-APR-1996 (first entry)
DE Potato citrate synthase.
KW Citrate synthase; inhibitor; increased storage capacity; potato;
KW antisense DNA.
OS Solanum tuberosum.
PN DE4408629-A1.
PD 14-SEP-1995.
PF 09-MAR-1994; 408629.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PR (AGRE) ROECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landtschutze V, Muller-Roeber B, Landtschutze V;
DR WPI; 95-321536/42.
N-PSDB; T03410.
PT Inhibiting citrate synthase (CS) activity in plants - to inhibit
PT flower formation and improve storage capacity, e.g. in potatoes,
PT also new CS sense and anti-sense DNA sequences
PS Claim 5; Page 15-19; 35pp; German.
CC The potato citrate synthase (CS) gene or cDNA sequence encoding CS
CC (T03410) may be used to produce antisense CS sequences. CS DNA
CC sequences are useful for altering CS activity in plants. Antisense
CC CS sequences can be used to inhibit CS expression in plants and has
CC the effect of inhibiting flower formation and by doing so improves
CC the plant's storage capacity. This is partic. useful in crop plants
CC of any kind but esp. useful in potatoes. In addition to altering CS
CC activity the DNA sequences can also be used to identify similar
CC sequences in the genomes of other plants and in the production of
CC transgenic plants with altered CS activity. This sequence
CC represents the amino acid sequence of potato citrate synthase.
SQ Sequence 471 AA;

Query Match 100.0%; Score 3466; DB 15; Length 471,
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mvfyrsvllsklsravqgnsvrwlqvtssgldlrsvlqelpeqqrlkkik 60

```

QY 1 MVFYPSVLSKLPRAVQSNVSRVWLVQVTSGLDLSELVQELIPQQDLKKIK 60
Db 61 sdmksgnltvdmvlgmrgmtgllwkpghyldpdegirfrglsipeccqkvlpaakpgge 120
QY 61 sdmksgnltvdmvlgmrgmtgllwkpghyldpdegirfrglsipeccqkvlpaakpgge 120
Db 121 pipeglwllltgkpskeqvnsivsgiaesgiisliimtyttidalpvtahpmtqfatgv 180
QY 121 PLPEGLLWLLLTGKPSKEQVNSIVSGIAESGIISLIIMTYTTIDALPVTAHPTQFATGV 180
Db 181 malqvsefagkayekghkskwyepedsmliaqvplvaayvyrmykngdipkdes 240
QY 181 PLPEGLLWLLLTGKPSKEQVNSIVSGIAESGIISLIIMTYTTIDALPVTAHPTQFATGV 180
Db 181 malqvsefagkayekghkskwyepedsmliaqvplvaayvyrmykngdipkdes 240
QY 181 MALQVSEFQKAYEKGHKSKEYEPTEDSMNLIAQVPLVAAYVYRMYKNGDIPKDES 240
Db 241 ldvganfahmifsssemhellmrlvytlhsdhegnvsahthlvasalsdpvlsfaaa 300
QY 241 LDYGANFAHMLGFSSEMHELLMRLVYTIHSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 301 lnglagplhqlanqevllwiksvveecgeniskeqlkdyvktlmsgkvpvfgchgvlrk 360
QY 301 LNLGAPLHGLANQEVLLWIKSVVEECGENISKEQLKDYVWKTLSNGKVPVFGHGVLRK 360
Db 361 tvprytcqrefamkhlpedplfqlvsklyevflfqlnklaklkwpmnvdaahsgvllnyg 420
QY 361 TVPRYTCQREFAMKHLPEPLFQLVSKLYEVFLFQLNKLAKLKPMPNVDAHSGVLLNYG 420
Db 421 ltearytytlfgvsralgicqlidwralglpkrpsvtmwlenqckka 471
QY 421 LTEARYTYTLFGVSRALGICSLIDWDRALGLPRLPKSVTMEWLENQCKKA 471

RESULT 2
ID R82838 standard; Protein: 471 AA.
AC R82838;
DT 25-JAN-1996 (first entry)
DE Potato citrate synthase.
KW Citrate synthase; flower formation; tuber storage.
OS Solanum tuberosum.
PN W09524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE ) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR N-PSDB: T04201.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Disclosure; Page 53-56; 87pp; English.
CC To identify a cDNA from potato which codes for citrate synthase, a
CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
CC which are complementary to the 5' or 3' end of the coding region of
CC A. thaliana cDNA for citrate synthase. The oligos additionally
CC introduce BamHI cleavage sites at both ends of the amplified cDNA
CC fragment. a cDNA library was prepd. from potato leaves and screened
CC with A. thaliana citrate synthase cDNA. Positive clones were
CC purified and sequenced. The nt sequence is given in T04199.
SQ Sequence 471 AA;

Query Match 100.0%; Score 3466; DB 14; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mvfyrsvllsklravqsnvsvrwlqvtsstgldlrselvgelipeqgdrllkik 60
QY 1 MVFYRSVLLSKLRSAVQSNVSRVWLVQVTSGLDLSELVQELIPQQDLKKIK 60
Db 61 sdmksgnltvdmvlgmrgmtgllwkpghyldpdegirfrglsipeccqkvlpaakpgge 120

```

```

QY 61 sdmksgnltvdmvlgmrgmtgllwkpghyldpdegirfrglsipeccqkvlpaakpgge 120
Db 121 pipeglwllltgkpskeqvnsivsgiaesgiisliimtyttidalpvtahpmtqfatgv 180
QY 121 PLPEGLLWLLLTGKPSKEQVNSIVSGIAESGIISLIIMTYTTIDALPVTAHPTQFATGV 180
Db 181 malqvsefagkayekghkskwyepedsmliaqvplvaayvyrmykngdipkdes 240
QY 181 MALQVSEFQKAYEKGHKSKEYEPTEDSMNLIAQVPLVAAYVYRMYKNGDIPKDES 240
Db 241 ldvganfahmifsssemhellmrlvytlhsdhegnvsahthlvasalsdpvlsfaaa 300
QY 241 LDYGANFAHMLGFSSEMHELLMRLVYTIHSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 301 lnglagplhqlanqevllwiksvveecgeniskeqlkdyvktlmsgkvpvfgchgvlrk 360
QY 301 LNLGAPLHGLANQEVLLWIKSVVEECGENISKEQLKDYVWKTLSNGKVPVFGHGVLRK 360
Db 361 tvprytcqrefamkhlpedplfqlvsklyevflfqlnklaklkwpmnvdaahsgvllnyg 420
QY 361 TVPRYTCQREFAMKHLPEPLFQLVSKLYEVFLFQLNKLAKLKPMPNVDAHSGVLLNYG 420
Db 421 ltearytytlfgvsralgicqlidwralglpkrpsvtmwlenqckka 471
QY 421 LTEARYTYTLFGVSRALGICSLIDWDRALGLPRLPKSVTMEWLENQCKKA 471

RESULT 3
ID R82840 standard; Protein: 469 AA.
AC R82840;
DT 25-JAN-1996 (first entry)
DE Tobacco citrate synthase.
KW Citrate synthase; flower formation.
OS Nicotiana tabacum
PN W09524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE ) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR N-PSDB: T04201.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Disclosure; Page 60-63; 87pp; English.
CC To identify a cDNA from tobacco which
CC synthase, a cDNA bank of leaf tissue from tobacco was prepd.
CC plaques of this cDNA bank were screened using a radioactive DNA
CC probe which comprises Solanum tuberosum citrate synthase cDNA
CC (T04199). One of the clones was sequenced. The nt. sequence is
CC given in T04201.
SQ Sequence 469 AA;

Query Match 89.4%; Score 3098; DB 14; Length 469;
Best Local Similarity 89.6%; Pred. No. 1.27e-270;
Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4;

Db 1 mvfyrgvllsklravqsnvsvrwlqvtsstgldlrselvgelipeqgdrllkik 59
QY 1 MVFYRSVLLSKLRSAVQSNVSRVWLVQVTSGLDLSELVQELIPQQDLKKIK 60
Db 60 sehkgvqlgnltvdmvlgmrgmtgllwetslldpdegirfrglsipeccqkvlpaakpgg 119
QY 61 sdmksgnltvdmvlgmrgmtgllwkpghyldpdegirfrglsipeccqkvlpaakpgg 119
Db 120 epelpgllwllltgkpskeqvnsivsgiaesgiisliimtyttidalpvtahpmtqfatg 178
QY 120 EPLPGLLWLLLTGKPSKEQVNSIVSGIAESGIISLIIMTYTTIDALPVTAHPTQFATG 179

```


Db 179 vma1qvsefgekayekighsksklweptyedsmsliaqplvaayvymkngntipkdd 238
 QY 180 VMA1QVSEFGEKAYEKGHKSKEYEPTIEDSMNLIAQPLVAAYVYRMKNGDTIPKDE 239
 Db 239 sldyganfahmlgfsdssdmhel-mklyvtihsdhegnvsahtghlvasalsdpylsfaa 297
 QY 240 SLDYGANFAHMLGFSSESSEMHELLMRLXYTIHSDHEGNVSAHTGHVVASALSOPYLSFAA 299
 Db 298 alnglaglqhlanqevllwksvveecgeniskeqlkdavaktlksqkvvpqfghv1r 357
 QY 300 ALNGLAGPLHGLANQEVLLWKSVEECGENISKEQLKDVWTKLNSGKVVPGFGHGVLR 359
 Db 358 ktdbryctqrefalkhlpdpqlfqlvaklyevflqnlaklnpwpnvvdahsgvllnyy 417
 QY 360 KTVPRYTCQREFAMKHLPEDFQLVSKLYEVFLFLQNLAKLKPWPNVDHSGVLLNYY 419
 Db 418 gltearyytlfvgvsralgicsqlidwralglplerpksvtmewlenhckka 469
 QY 420 GLTEARYYTLFVGSRALGICSQLIDWALGLPLERPCKSVTMEWLENQCKKA 471

RESULT 4
 ID R82839 standard; Protein; 437 AA.
 AC R82839;
 DT 25-JAN-1996 (first entry)
 DE Sugar beet citrate synthase.
 KW Citrate synthase; flower formation.
 OS Beta vulgaris strain Zuchtlinie 5S 0026
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) ROECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutz V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR N-PSDB; T04200.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Disclosure; Page 57-60; 87pp; English.
 CC To identify a cDNA from sugar beet which codes for citrate
 CC synthase, a cDNA bank of leaf tissue from sugar beet was prepd.
 CC Plaques of this cDNA bank were screened using radioactive DNA
 CC probes which comprise a mixture of Solanum tuberosum citrate
 CC synthase cDNA (T04199) and Nicotian tabacum citrate synthase cDNA
 CC (see T04201). One of the clones was sequenced. the nt. sequence is
 CC given in T04200.
 SQ Sequence 437 AA;

Query Match 76.1%; Score 2637; DB 14; Length 437;
 Best Local Similarity 81.3%; Pred. No. 8,02e-228;
 Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;

Db 1 ssndlrsl-gelipeqgerlkkikkfsgfignvdmvgmrgmtglwetslld 59
 QY 35 SSGLDLRSELQELPEQQRLLKIKSDMKG-SIGNITVDMVLGMRGMLLWKKPHYLD 93
 Db 60 peegirfrgrfipccqklpaasagaeppegllwllltkvpkskevdsalsadlrkas 119
 QY 94 PDEGIRFGLSIPECQKYLPAKFGGEPLEGLLWLLLTGKVPKSEQVNSIVSIAESGI 153
 Db 120 ipdhv-yktidalpitaqmtqfctymalqtrefekayekighkskfweptyedcsl 178
 QY 154 ISLIIMYTTIDALPVTAHPMTQFATGVALQVSEFOKAYEKGHKSKEYEPTIEDSMNL 213
 Db 179 iaqvpvaayvymkngvpiiddidygnfnahmlgfdspqmlsl-mrlyvtihsdh 237
 QY 214 IAQVPLVAAYVYRMKNGDTIPKDES LDYGANFAHMLGFSSESSEMHELLMRLXYTIHSDH 273
 Db 238 eggnvsahtghlvspisdpylsfaa1nglaglqhlanqevllwksvveecgenist 297
 QY 239 eggnvsahtghlvspisdpylsfaa1nglaglqhlanqevllwksvveecgenist 297

QY 274 EGGNVSAHTGHVVASALSOPYLSFAAALNGLAGPLHGLANQEVLLWKSVEECGENISK 333
 Db 298 eqlkvvsktlnsgkvpgfqlgvlrkttdprytcqrefalkhlpdpfqlvsklyevvp 357
 QY 334 EQLKDVWTKLNSGKVPFGFGHGVLRKTVPRYTCQREFAMKHLPEDFQLVSKLYEVFL 393
 Db 358 pillelkgvknppnvvdahsgvllnhgyltearyytlfvgvsralgicsqlidwralglp 417
 QY 394 LFLQNLAKLKPWPNVDHSGVLLNYYGLTEARYYTLFVGSRALGICSQLIDWALGLP 452
 Db 418 lerpsvtmewlekfckr 435
 QY 453 LERPCKSVTMEWLENQCK 470

RESULT 5
 ID R14357 standard; Protein; 436 AA.
 AC R14357;
 DT 23-JAN-1992 (first entry)
 DE Citrate synthase encoded by aara.
 KW Acetic acid resistance.
 OS Acetobacter sp.
 PN J03219878-A.
 PD 27-SEP-1991.
 PF 05-FEB-1990; 024395.
 PR 15-FEB-1989; JP-033776.
 PR 05-FEB-1990; JP-024395.
 PA (NAKA-) NAKANO SUMISE KK.
 DR WPI; 91-329112/45.
 DR N-PSDB; Q14367.
 PT Acetic acid resistant gene with in plasmid and transformed
 PT Acetobacter - improves yield of acetic acid fermentation.
 PS Disclosure; Fig 5; 12pp; Japanese.
 CC The aara gene encodes the citrate synthase and is part of an
 CC acetic acid resistance operon comprising aara, aaraB and aaraC.
 CC The DNA can be used to prepare an acetic acid resistant strain of
 CC bacteria for use in acetic acid fermentation.
 CC See also R14358 and 59.
 SQ Sequence 436 AA;

Query Match 8.9%; Score 310; DB 3; Length 436;
 Best Local Similarity 30.5%; Pred. No. 1.16e-15;
 Matches 100; Conservative 85; Mismatches 11; Indels 32; Gaps 28;

Db 91 eevivillngelpnkaqdytftntlnhtllheqir-nfngfrdrdahpmallogtval 149
 QY 124 EELLWLLLTGKVPKSEQVNSIVSGIAESGIISLIIMYTTIDALPVTAHPMTQFATGVAL 183
 Db 150 sa-f-ydpdandiaip-an-rdia---amrlakiptiaawayk--ytggeafiyprnd-l 199
 QY 184 QVQSEFOKAYEKGHKSKEYEPTIEDSMNLIAQVPLVAAYVYRMKNGDT-I-PKDESL 241
 Db 200 nyaenflmmfarmsepkyvnpvlaramnrillhadheq-naststvrslagstangfa 258
 QY 242 DYGANFAHML-G-FSSS-EMHELLMR-LY-VTI-HSDHEGNGVSAHTGHVVASALSOPYL 295
 Db 259 claaqialwghpahnaganeavik-mlarigk-kenip-afiaq-v-kdknsgvklmgfgh 313
 QY 296 SFAAALNGLAGPLHGLANQEVLLWKSVEECGENISKEQLKDYVWTKLNSGKVVPGFGH 355
 Db 314 rvykfndprakimqqtchevtelgikddplldlavel-ekialsdddyfqrklypnvdf 372
 QY 356 GVLKRTVPRYTC-QRE-FA-MKHLR--EDPLFOLVSKLYEVFLFLQNLAKLKPWPNVDA 410
 Db 373 ysgillkamgipstsmf-tvlfavattg 399
 QY 411 HSGVLLNYYGLTEARYYTLFVGSRALG 438

RESULT 6
 ID W69756 standard; Protein; 754 AA.
 AC W69756;
 DT 23-NOV-1998 (first entry)

DE Acetobacter xylinum bcsA gene product.
 KW Acetobacter xylinum: sucrofermentans; cellulose synthetase complex;
 KW bcsA; bcsB; bcsC; bcsD; CMCase; beta-glucosidase; enzyme; cellulose;
 KW microorganism.
 OS Acetobacter xylinum.
 FH Key Location/Qualifiers
 FT Misc_difference 503 /label= unspecified
 FT /note= "encoded by GAG"
 FT Misc_difference 506 /label= unspecified
 FT /note= "encoded by GAG"
 FT Misc_difference 573 /label= unspecified
 FT /note= "encoded by AAG"
 PN WO9839455-A1.
 PD 11-SEP-1998.
 PF 09-OCT-1997; J03633.
 PR 04-MAR-1997; JP-063927.
 PA (BIOP-) BIO-POLYMER RES CO LTD.
 PI Hayashi T, Tahara N, Tonouchi N, Tsuchida T, Yano H,
 PI Yoshinaga F.
 DR WPI; 98-495854/42.
 DR N-PSDB; V52831.
 PT Gene encoding Acetobacter xylinum cellulose synthetase complex -
 PT containing a group of genes including those for conventional and
 PT novel α -glucosidases
 PS Claim 2: Page 34: 50pp; Japanese.
 CC This represents the amino acid sequence of a Acetobacter xylinum
 CC subspecies sucrofermentans bcsA gene product. The invention provides
 CC a gene encoding a Acetobacter xylinum subspecies sucrofermentans derived
 CC cellulose synthetase complex-produced protein. The gene sequence
 CC represents bcsA, bcsB, bcsC or bcsD, CMCase and a beta-glucosidase
 CC encoding gene. The novel gene and the enzyme participate in the synthesis
 CC of cellulose by microorganisms. Cells transformed with the genes may be
 CC used in the production of cellulose.
 SQ Sequence 754 AA;
 Query Match 3.8%; Score 130; DB 35; Length 754;
 Best Local Similarity 19.4%; Pred. No. 2,91e-01;
 Matches 27; Conservative 48; Mismatches 55; Indels 9; Gaps 8;
 Db 420 plaffagqniiaaplavaayalpbmhfshiata-akv-nkg-wrysfwsevyetmal.f 476
 QY 155 SLIIMYTTIDALPVTAPHTQFATGVMAQVQSEFOKAYEKGIHKSKEYEPTYEDSMNL- 213
 Db 477 lvrvtiwllfpgskgnvtekgvixeexfldiqatynii-fatimmgglligifeli.v 535
 QY 214 IAQVPLVAA-VYVRRMYKNGDT--IPKDESLDYGANFAHMLGFSSEMHELLMRLYVTH 270
 Db 536 rfngl-dviarnayllnca 553
 QY 271 SDHEGNGVSAHTGHLVASA 289
 RESULT 7
 ID R47340 standard; Protein; 220 AA.
 AC R47340;
 DT 01-JUL-1994 (first entry)
 DE Fragment of chromaffin granule amine transporter protein.
 KW Vesicle membrane transport protein; gene therapy; screening;
 KW Parkinsons disease; neurotoxin; identification; detection;
 KW antibody; probe;chromaffin granule amine transporter protein.
 OS Rattus rattus.
 PN WO9325699-A.
 PD 23-DEC-1993.
 PF 11-JUN-1993; U05704.
 PR 11-JUN-1992; US-899074.
 PR 30-JUL-1992; US-923096.
 PA (REGC) UNIV CALIFORNIA.
 PI Edwards RH;
 DR WPI; 94-007556/01.
 PT New mammalian vesicle membrane transport protein - and corresp.

PT DNA, vectors, transformed cells and antibodies, for diagnosis and
 PT treatment of neurological disorders, e.g. Parkinson disease
 PS Example 2: Page 114-115; 181pp; English.
 CC The cDNA encoding the chromaffin granule amine transporter protein
 CC is useful in gene therapy and as a probe for detecting genomic
 CC sequences. The protein is used for screening cytotoxic compounds
 CC implicated in Parkinsons disease, diseases associated with activity
 CC of neurotoxins or psychiatric disorders and to identify compounds
 CC which selectively inhibit or activate its action. Antibodies raised
 CC against this protein are useful as immunoassay reagents for
 CC detecting the protein and as affinity reagents for purification.
 CC This fragment of the transporter protein shares homology with the
 CC N-terminal domains of the methlenomycin; tetracycline and multi-drug
 CC resistance transporter proteins.
 SQ Sequence 220 AA;
 Query Match 3.3%; Score 113; DB 9; Length 220;
 Best Local Similarity 20.2%; Pred. No. 4.63e+00;
 Matches 22; Conservative 41; Mismatches 41; Indels 5; Gaps 5;
 Db 24 dnmlltvvpiptflyatefkdsnsllhrqpsvssgeenvriillfaskalmollvnpf 83
 QY 209 DSMNLIAQVPLVAAVYVRRMYKNGDT-IPKDESLDYGANFAHM-LGFSSSEMHELLMRLY 266
 Db 84 vgpitnriqyhipmfvgfmfmf-1stlmfatsgtviallvartlqqigs 131
 QY 267 VTIIHSDHEGNGVSAHTGHLVASALSDPYLSPFAALNGL-AG-PLHGLAN 313
 RESULT 8
 ID R45000 standard; Protein; 754 AA.
 AC R45000;
 DT 13-JUN-1994 (first entry)
 DE Cellulose synthase operon, gene A product.
 KW Bacterial; cellulose synthase; operon; gene A; gene B; gene C;
 KW gene D; transcription vector; recombinant microorganism;
 KW cellulose synthesis.
 OS Acetobacter xylinum.
 PN US5268274-A.
 PD 07-DEC-1993.
 PF 12-APR-1989; 337194.
 PR 04-APR-1990; WO-U01811.
 PR 09-APR-1990; IL-094053.
 PR 10-APR-1990; CA-014264.
 PR 11-APR-1990; IE-001317.
 PR 12-APR-1990; NZ-233312.
 PA (CETU) CETUS CORP.
 PI Ben-Bassat A, Ben-Ziman M, Calhoon RD, Fear AL, Gelfand DH;
 PI Meade JH, Tal R, Wong H;
 DR WPI; 93-404004/50.
 DR N-PSDB; 053522.
 PT Contiguous nucleic acid sequences - encoding bacterial cellulose
 PT synthase
 PS Claim 50; Fig 1; 79pp; English.
 CC The sequences given in R45000-03 represent the proteins encoded by
 CC the bacterial cellulose synthase operon. The sequence given in
 CC R45004 is the beginning of an open reading frame overlapping the end
 CC of this operon. The bacterial cellulose synthase operon contains
 CC four genes, genes A-D. The operon sequence may be used in a
 CC transcription vector for the expression of the cellulose synthase
 CC operon to increase cellulose production in a recombinant microorganism.
 CC This system may be used as an important tool for exploring mechanisms
 CC of cellulose synthesis and for enhancing production of cellulose.
 SQ Sequence 754 AA;
 Query Match 3.3%; Score 115; DB 8; Length 754;
 Best Local Similarity 18.1%; Pred. No. 3.37e+00;
 Matches 23; Conservative 46; Mismatches 49; Indels 9; Gaps 8;
 Db 420 plaffagqniiaasplaviayalpbmhfshiata-akv-nkg-wrysfwsevyetmal.f 476
 QY 155 SLIIMYTTIDALPVTAPHTQFATGVMAQVQSEFOKAYEKGIHKSKEYEPTYEDSMNL- 213

PD	29-OCT-1997.		
PF	27-MAR-1997.	U05223.	
PR	06-DEC-1996;	US-761318.	
PR	29-MAR-1996;	US-625811.	
PR	02-APR-1996;	US-758731.	
PR	25-OCT-1995;	US-736905.	
PR	28-OCT-1995;	US-738859.	
PA	(ASTR) ASTRA AB.		
PI	Alm RA, Smith D.		
DR	WPI; 97-503122/46.		
DR	N-PSDB; V24983.		
PT	Helicobacter pylori nucleic acid sequences and encoded		
PT	polypeptide(s) - useful in vaccines to treat or prevent		
PI	<i>H. pylori</i>		

RESULT	14				
ID	R97244	standard; Protein; 4473 AA.			
AC	R97244;				
DC	07-JAN-1997	(first entry)			
DE	Virulence gene cluster polypeptide product.				
DE	Mutant; adaptation; virulence factor; identification; screening;				
KW	vaccine; drugs; infection; treatment.				
KW	Salmonella typhimurium.				
OS	Salmonella typhimurium.				
FT	Key	Location/Qualifiers			
FT	Region				
FT	FT	/note= 'All x's in this sequence correspond to			
FT	FT	termination codons in the virulence gene			
FT	FT	cluster sequence given in T09224."			
PN	W09617951-A2.				
PD	13-JUN-1996.				
PF	11-DEC-1995; G02875.				
PF	09-DEC-1994; GB-024921.				
PR	31-JAN-1995; GB-001881.				
PR	05-MAY-1995; GB-009239.				
PA	(RPMS-) RPMS TECHNOLOGY LTD.				
PI	Holden DW;				
DR	WPI; 96-287194/29.				
DR	N-PSDB; T09224.				
PT	Identifying virulence genes in microorganisms - by introducing				
PT	mutants with insertion inactivated genes into environment and				
PT	retrieval and analysis of mutants				
PS	Claim 51; Figure 11; 13pp; English.				
CC	A method for identifying a microorganism having a reduced adaptation				
CC	to a particular environment comprising the steps of: (1) providing a				
CC	plurality of microorganisms each of which is independently mutated by				
CC	the insertional inactivation of a gene with a nucleic acid comprising				
CC	a unique marker sequence so that each mutant contains a different				
CC	marker sequence, or clones of the said microorganism; (2) providing				
CC	individually a stored sample of each mutant produced by step (1) and				
CC	providing individually stored nucleic acid comprising the unique				
CC	marker sequence from each individual mutant; (3) introducing a				
CC	plurality of mutants produced by step (1) into the said particular				
CC	environment and allowing those microorganisms which are able to do so				
CC	to grow in the said environment; (4) retrieving microorganisms from				
CC	the said environment or a selected part thereof and isolating the				
CC	nucleic acid from the retrieved microorganisms; (5) comparing any				
CC	marker sequences in the nucleic acid isolated in step (4) to the				
CC	unique marker sequence of each individual mutant stored as in step				
CC	(2); and (6) selecting an individual mutant which does not contain any				
CC	of the marker sequences as isolated in step (4). The products and				
CC	methods can be used for identifying virulence genes in microorganisms.				
CC	The mutant microorganisms can be used in vaccines or to screen for				
CC	drugs which reduce virulence or compounds useful for preventing,				
CC	ameliorating or treating infections in animals or plants.				
SQ	Sequence 4473 AA;				
Query Match 2.9%; Score 100; DB 19; Length 4473;					
Best Local Similarity 22.9%; Pred. No. 3 50e+01;					
Matches 19; Conservative 25; Mismatches 35; Indels 4; Gaps 4;					
Db	2872	mtixxrscrsherysms-lpdspl-qllgilfililpl-ilvmqtsfklavv-silr	2928		
QY	358	LRTVPRY-TCQREFAMKHLPEDLPQLSKLYEVLFLQLAKLPWPVNDHSGVLL	416		
Db	2929	nalgiqgvppnialglavlsl	2951		
QY	417	NYGLTEARYTYVLFVGSRALGI	439		
RESULT	15				
ID	W26588	standard; Protein; 355 AA.			
AC	W26588;				
DC	21-JAN-1998	(first entry)			
DE	Human MIP-1 alpha/RANTES				
DE	Macrophage inflammatory protein-1 alpha; MIP-1 alpha;				
KW	reduced upon activation normal T expressed and secreted; RANTES;				



Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	3466	100.0	471	2	T44316	ethanolamine ammonia-	0.00e+00
2	2820	81.4	474	2	S03390	probable citrate (sl)	0.00e+00
3	2305	66.5	472	1	YKMUM	citrate (sl)-synthase	0.00e+00
4	1921	55.4	464	1	YPGP	citrate (sl)-synthase	0.00e+00
5	1722	49.7	468	2	S42370	citrate (sl)-synthase	0.00e+00
6	1663	48.0	479	1	YKBY	citrate (sl)-synthase	9.09e-29
7	1606	46.3	469	2	S15563	citrate (sl)-synthase	2.84e-28
8	1575	45.4	460	1	YKBYC	citrate (sl)-synthase	4.46e-27
9	1329	38.3	462	2	J55625	14 nm filament protei	8.43e-22
10	1168	33.7	486	2	S52814	citrate (sl)-synthase	9.90e-19
11	316	9.1	437	2	I40717	citrate (sl)-synthase	3.00e-33
12	307	8.9	436	1	YKPC	citrate (sl)-synthase	1.13e-31
13	272	7.8	428	1	YKPSCA	citrate (sl)-synthase	1.26e-25
14	266	7.7	516	2	S33007	citrate synthase - cu	1.33e-24
15	263	7.6	397	2	S74344	citrate synthase gltA	4.28e-24
16	261	7.5	372	2	C69417	citrate synthase (cit	9.34e-24
17	254	7.3	424	2	J39506	citrate (sl)-synthase	1.41e-22
18	254	7.3	430	2	Q01392	citrate (sl)-synthase	1.41e-22
19	249	7.2	373	2	A43936	citrate synthase - Ba	9.75e-22
20	251	7.2	431	2	E70782	probable gltA2 protei	4.54e-22
21	245	7.1	436	1	YKRECP	citrate (sl)-synthase	4.54e-22
22	242	7.0	372	2	E59568	citrate synthase III	1.43e-20
23	238	6.9	431	2	I40044	citrate (sl)-synthase	6.59e-20

```

QY 361 TVPRYTQREFAMKHLPEDFQLVSKLYEVFLFLQNLAKLPWPNVDAHSGVLLNYG 420
|||||
Db 421 LTEARYTYTVLFGVSRALGICSQLINDRALGLPLERPKSVTMEWLENQCKKA 471
|||||
QY 421 LTEARYTYTVLFGVSRALGICSQLINDRALGLPLERPKSVTMEWLENQCKKA 471
|||||

RESULT 2 T02390 #type complete
ENTRY probable citrate (si)-synthase (EC 4.1.3.7) - Arabidopsis
TITLE thaliana
ALTERNATE_NAMES F411.16 protein
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02390
REFERENCE 214177
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandson, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.F.; Venter,
J.C.
#submission submitted to the EMBL Data Library, May 1998
#description Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
#accession T02390
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-474 ##label ROU
##cross-references EMBL:AC004521; NID:g3128166; PID:g3128180
GENETICS
#map_position 2
#introns 17/3; 39/3; 53/3; 75/3; 97/3; 135/3; 157/1; 184/3; 201/2;
227/2; 270/2; 284/3; 313/3; 347/3; 363/3; 400/3; 425/2;
442/3
#note F411.16
KEYWORDS carbon-carbon lyase; oxo-acid-lyase; tricarboxylic acid cycle
SUMMARY #length 474 #molecular-weight 52782 #checksum 6960

Query Match 81.4%; Score 2820; DB 2; Length 474;
Best Local Similarity 79.1%; Pred. No. 0.00e+00;
Matches 375; Conservative 55; Mismatches 38; Indels 6; Gaps 6;

Db 1 MVFFRSVSAFTRLRSRVQGOSSLSNSVRWIOQSTDLKSQL-QELIPEQDRLKKL 59
|||||
QY 1 MVFYSVSLSLKRSRA-VQOSNVSNVSRVWLQVOTSSGLDLRSELVQELIPEQDRLKKI 59
|||||
Db 60 KSEHGKVLQGNITVDVIGCMRGMTGLLWETSLLDPEGIRFGLSIPECQKVLPTAOSG 119
|||||
QY 60 KSDM-KGSIGNITVDVIGCMRGMTGLLWKPHYLDPEGIRFGLSIPECQKVLPAAPKG 118
|||||
Db 120 AEPPLGELLWLLTGKVPKSEQVLSKDLANRAAVPDYV-YNAIDALPSTAHPTQFAS 178
|||||
QY 119 GEPLPEGLLWLLTGKVPKSEQVNSIVSGIASGIIISLIIMTTIDALPVTAHPTQFAT 178
|||||
Db 179 GVMALQVQSEFQKAYENGIIHKSKEFWEPTEDCNLIARVPVVAAYVYRMYKNGDSIPS 238
|||||
QY 179 GVMALQVQSEFQKAYEKGIHKSKEFWEPTEDSMNLIAQVPLVAAVYRMYKNGDTIPKD 238
|||||
Db 239 KSLDYCANFSHMLGFDDEKVKEL-MRLYTIHSDHEGNGVNSAHTGLVGSALSDPYLSFA 297
|||||
QY 239 ESLDYGANFAHMLGFSSEMHELLMRLYVTIHSDEHGGNVSAHTGHLVASALSDPYLSFA 298
|||||
Db 298 AALNGLAGPLHGLANQEVLLWIKSVVEECGDISKEQLKEYVWKTLNSGKVIPIGYGHVL 357
|||||
QY 299 AALNGLAGPLHGLANQEVLLWIKSVVEECGENISKEQLKDYVWKTLNSGKVPFGHGV 358
|||||
Db 358 RNTDPRYVQCFQKALHLPDDPLFQVLSKLYEVPVPTTELGVKNPNPNVDAHSGVLLN 417
|||||
QY 359 RKTVPYTCQREFAMKHLPEDFQLVSKLYEVFLFLQNLAKLPWPNVDAHSGVLLN 417
|||||
Db 418 HYGLTARYTYTVLFGVSRALGICSQLINDRALGLALERPKSVTMDWLEAHCKKA 471
|||||

```

```

QY 418 YGGLTEARYTYTVLFGVSRALGICSQLINDRALGLPLERPKSVTMEWLENQCKKA 471
|||||
RESULT 3 YKMUM #type complete
ENTRY citrate (si)-synthase (EC 4.1.3.7) precursor, mitochondrial
TITLE Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
ACCESSIONS JA0149; S07004
REFERENCE JA0149
#authors Unger, E.A.; Hand, J.M.; Cashmore, A.R.; Vasconcelos, A.C.
#journal Plant Mol. Biol. (1989) 13:411-418
#title Isolation of a cDNA encoding mitochondrial citrate synthase
from Arabidopsis thaliana.
#cross-references MUID:91370823
#accession JA0149
##molecule_type mRNA
##residues 1-472 ##label UNG
##cross-references EMBL:X17528; NID:g11243; PID:e1188578; PID:g2652924
##note the sequence from Fig. 2 is inconsistent with that from
Fig. 1 in having 47-Gly, 125-Trp, 141-Leu, 151-Ser,
185-Asn, 187-Asn, 191-Asn, 348-Leu, an additional Gly
after 89-Trp, an additional Val after 393-Cys and two
additional Arg after 228-Arg, in lacking 114-Leu and
residues 233 to 241
GENETICS
#genome nuclear
#complex homodimer
#classification #superfamily citrate (si)-synthase
#keywords carbon-carbon lyase; homodimer; mitochondrion;
oxo-acid-lyase, tricarboxylic acid cycle
FEATURE
343-375 #region acetyl-CoA binding #status predicted\
309,355,407 #active_site His, His, Asp #status predicted
SUMMARY #length 472 #molecular-weight 52941 #checksum 9626

Query Match 66.5%; Score 2305; DB 1; Length 472;
Best Local Similarity 68.9%; Pred. No. 0.00e+00;
Matches 328; Conservative 60; Mismatches 76; Indels 12; Gaps 12;

Db 1 MVFFRSVSAFTRLRSRVQGOSSLSNSVRWIOQSTDLKSQL-QELIPEKHDLKKL 59
|||||
QY 1 MVFYSVSLSLKRSRA-VQOSNVSNVSRVWLQVOTSSGLDLRSELVQELIPEQDRLKKI 59
|||||
Db 60 KSEHGKVLQGNITVDVIGCMRGMTGLLWETSLLDPEEVALGDCRLPECQKALLPTAOS 119
|||||
QY 60 KSDM-KGSIGNITVDVIGCMRGMTGLLWKPHYLDPEGIRFGLSIPECQK-VLPAAPK 117
|||||
Db 120 GGLNHYRRSFVASLWKKGTI-AKSKLKH-CRKTWNAAYSDYV-YNAIDALPSTAHPTQ 176
|||||
QY 118 GG-EPLPEGLL-WLLTGKVPKSEQVNSIVSGIASGIIISLIIMTTIDALPVTAHPTQ 175
|||||
Db 177 FASGVMALQVQSEFQKAYENGIIHKSKEFWEPTEDCNLIARVPVVAAYVYRMYKNGDSI 236
|||||
QY 176 FATGVMALQVQSEFQKAYENGIIHKSKEFWEPTEDSMNLIAQVPLVAAVYRMYKNGDTI 235
|||||
Db 237 PSKSLDYCANFSHMLGFDDEKVKEL-MRLTSPSTVHDEGNGVNSAHTGLVGSALSDPYL 295
|||||
QY 236 PKDESLDYCANFAHMLGFSSEMHELLMRLYVTIHSDEHGGNVSAHTGHLVASALSDPYL 295
|||||
Db 296 SFAAALNGLAGPLHGLANQEVLLWIKSVVEECGDISKEQLKEYVWKTLNSGKVIPIGYGH 355
|||||
QY 296 SFAAALNGLAGPLHGLANQEVLLWIKSVVEECGENISKEQLKDYVWKTLNSGKVPFGH 355
|||||
Db 356 GVLNTPRYVQCFQKALHLPDDPLFQVLSKLYEVPVPTTELGVKNPNPNVDAHSGVLL 413
|||||
QY 356 GVLNTPRYVQCFQKALHLPDDPLFQVLSKLYEVPVPTTELGVKNPNPNVDAHSGVLL 415
|||||
Db 414 LNHYGLTARYTYTVLFGVSRALGICSQLINDRALGLALERPKSVTMDWLEAHCKKA 469
|||||

```



```

QY 416 LNYGLTEARYTYLVFGVSRALGICSLIWDRLGLPLERPKSVTMEWLENCKKA 471

RESULT 4 YKPG #type complete
ENTRY citrate (si)-synthase (EC 4.1.3.7) precursor - pig
TITLE #formal_name Sus scrofa domestica #common_name domestic pig
ORGANISM 15-Oct-1982 #sequence_revision 30-Sep-1992 #text_change
DATE 18-Sep-1998

ACCESSIONS A29965; A01109; A61347
REFERENCE A29966
#authors Evans, C.T.; Owens, D.D.; Simegi, B.; Kispal, G.; Srere, P.A.
#journal Biochemistry (1988) 27:4680-4686
#title Isolation, nucleotide sequence, and expression of a cDNA
#encoding pig citrate synthase.
#cross-references MUID:89000665
#accession A29966
#molecule_type mRNA
#residues 1-464 #label EVA
#cross-references EMBL:M21197; NID:g164418; PID:g164419
REFERENCE A30457
#authors Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Walsh, K.A.;
Titani, K.
#journal Biochemistry (1982) 21:2028-2036
#title Complete amino acid sequence of porcine heart citrate
synthase.
#cross-references MUID:82231993
#accession A01109
#molecule_type protein
#residues 28-464 #label BLO
REFERENCE A61347
#authors Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Wade, R.D.;
Ericsson, L.H.; Neurath, H.; Walsh, K.A.; Titani, K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:5381-5385
#title Primary structure of porcine heart citrate synthase.
#accession A61347
#molecule_type protein
#residues 28-464 #label BL2
REFERENCE A92884
#authors Remington, S.; Wiegand, G.; Huber, R.
#journal J. Mol. Biol. (1982) 158:111-152
#title Crystallographic refinement and atomic models of two
different forms of citrate synthase at 2.7 and 1.7
angstroms resolution.
#cross-references MUID:83010291
#contents annotation: X-ray crystallography, 2.7 and 1.7 angstroms
COMMENT Citrate (si)-synthase is found in nearly all cells capable of
oxidative metabolism. It catalyzes the condensation of
oxaloacetate and acetyl-CoA to form citrate in the tricarboxylic
acid cycle.
COMMENT It is synthesized in the cytoplasm but functions in the
mitochondrion of eukaryote cells.
COMMENT This molecule is a dimer of identical chains. Each dimer binds two
molecules of acetyl-CoA and two molecules of oxaloacetate at two
active sites.
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; homodimer; methylated amino acid;
mitochondrion; oxo-acid-lyase; tricarboxylic acid cycle
FEATURE
1-27 #domain transit peptide (mitochondrion) #status
#predicted #label SIG
301-347,402 #product citrate (si)-synthase #status experimental
#label NAT
395 #active_site His, His, Asp #status predicted\
#modified_site N6,N6,N6-trimethyllysine (Lys) #status
experimental
SUMMARY #length 464 #molecular-weight 51629 #checksum 3548
Query Match 55.4%; Score 1921; DB 1; Length 464;
Best Local Similarity 58.8%; Pred. No. 0.00e+00;
Matches 254; Conservative 86; Mismatches 86; Indels 6; Gaps 6;
Db 27 SASSTNLK-DILADLPKQCARIKTRFQGHGNTVVGQIVDDMMYGGMRCKGLVETSVL 85

```

```

QY 34 TSSGLDLPSELVQELIPEQCDPLKKIKSDMKGST-GNITVTVMLGMPGTMGLWKPHYL 92
Db 86 DPDEGIFPGYSTPECOQMLPKAKGGEPLPEGLFWLLVTGQIPTDEQVSWLSKENAKPA 145
QY 93 DPDEGIFRGLSIPCOQVLPAAKPGGEPLPEGLWLLTCKVPSKEQVNSIVSGIAESG 152
Db 146 ALPSHVY-TLMDNFTNLHPMSOLSAAITALNESNFARAYAEIGIHRKTKWELIYEDMD 204
QY 153 IISLIIMYTIDALPVTAHMTQFATGMALQVSEFOKAYEKIHKSKWEFTYEDSMN 212
Db 205 LIAKLPQVAAIKIYNLYREGSSIGSIDSKLDWSHNFNMILGYTDAQFTEL-MRLYLTIS 263
QY 213 LIAQVPLVAAVYVRRMYKNGDTI-PKDESLOYGANFAHMLGFSSEMHEHLMRLYVTIS 271
Db 264 DHEGGSYSAHTSHLVGSALSDPYLSFAAANGLAGPLHGLANQFVLYWLTQLOKEVGKV 323
QY 272 DHEGGSYSAHTGHVLSALSDPYLSFAAALNGLAGPLHGLANQFVLYWLTQVVEEGENI 331
Db 324 SDEKLRDIYIWNLTNSGRVPGYGHAVLRKTDPRYTCOREFALKHLPDHPMFKLVAOLYKI 383
QY 332 SKEOLKDYVVKTLNSGKVPFGHGVLRKTVPRYTCOREFAMKHLPEPLFQLVSKLYEV 391
Db 384 VPNVLLSQGKAKNPWPNDVDAHSGVLLQYGYGTEMNYYTVLVFGVSRALGVLAQIWSRALG 443
QY 392 FLFLQLAKLK-PWPNVDAHSGVLLNYYGLTEARYYTVLVFGVSRALGICSLIWDRLALG 450
Db 444 PFLERPKSMSTD 455
QY 451 LPLERPKSVTME 462

RESULT 5 S42370 #type complete
ENTRY citrate (si)-synthase (EC 4.1.3.7) precursor - Caenorhabditis
TITLE elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change
18-Sep-1998
ACCESSIONS S42370
REFERENCE S42368
#authors Smith, A.
#submission submitted to the EMBL Data Library, March 1994
#accession S42370
#status preliminary
#molecule_type DNA
#residues 1-468 #label SMI
#cross-references EMBL:Z30423; NID:g458479; PID:g458482
GENETICS
#introns 69/3; 202/3; 309/3
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; mitochondrion; oxo-acid-lyase
SUMMARY #length 468 #molecular-weight 51540 #checksum 2659
Query Match 49.7%; Score 1722; DB 2; Length 468;
Best Local Similarity 54.0%; Pred. No. 0.00e+00;
Matches 236; Conservative 84; Mismatches 109; Indels 8; Gaps 8;
Db 25 LSTSAEGSTNLKVELSKK-IPAHNAKVKSPRTEHSGTVVQNVNIDMIYGGMRCKGMVTE 83
QY 30 LOVOTSSGLDRSELVQELIPEQODRLKKIKSDMKGSIG-NITVDVWVGMRGTMGLLWK 88
Db 84 TSVLDPEGIFRFGYSIPECOQKLLPKAKGGEPLPEAIWLLCTGDVPSQAATAITKEW 143
QY 89 PHYLDPEGIFRGLSIPCOQVLPAAKPGGEPLPEGLWLLTGVPSKEQVNSIVSGI 148
Db 144 NARADLPETHV-RMLDNFPDNLHPMAQFIAIAALNNEKFAAGYARGVAKASWYFAYE 202
QY 149 AESGIISLIIMYTIDALPVTAHMTQFATGMALQVSEFOKAYEKIHKSKWEFTY 208
Db 203 DSMDLAKLPVAAIYYRNLVROGSVSVDPKDKWSANFSSMLGYDDPLFAEL-MRLYL 261
QY 209 DSMNLIAQVPLVAAVYVRRMYKNGDTIP-KDESLOYGANFAHMLGFSSEMHEHLMRLYV 267

```



```

QY 384 LVSKLVEYVLLFQNLAKLK-PWPNDVDSHGVLLNYGLTEARYTYTLFGVSRALGTCQ 442
      ||| :||| | : : | | | | | | | | | | | | | | | | : : | | | | | | | | | : |
Db 428 LITDRAIGASIERPKSYSTE 447
      ||| :||| : : | | | | | : |
QY 443 LIWDPALGLPEPKSYTME 462
      ||| :||| : : | | | | | : |

RESULT 9
ENTRY #type complete
TITLE 14-nm filament protein/citrate synthase (EC 4.1.3.-) -
      Tetrahymena thermophila (SGC5)
ALTERNATE_NAMES 49K filament-forming protein
ORGANISM #formal_name Tetrahymena thermophila
DATE 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change
ACCESSION JG5625
REFERENCE JG5625; JN0130
#authors Takeda, T.; Watanabe, Y.; Numata, O.
#journal Biochem Biophys Res Commun (1997) 237:205-210
#title Direct demonstration of the bifunctional property of
      Tetrahymena 14-nm filament protein/citrate synthase
      following expression of the gene in Escherichia coli.
#accession JG5625
#molecule_type mRNA
#residues 1-462 #label TAK
REFERENCE JN0130
#authors Numata, O.; Takemasa, T.; Takagi, I.; Hirono, M.; Hirano, H.;
      Chiba, J.; Watanabe, Y.
#journal Biochem Biophys Res Commun. (1991) 174:1028-1034
#title Tetrahymena 14-nm filament-forming protein has citrate
      synthase activity.
#cross-references GMD:91128358
#accession JN0130
#molecule_type mRNA
#residues 1-462 #label NUM
#cross-references GB:D90117; NID:g217406; PID:d1014848; PID:g1688046
#note part of this sequence, including the amino end of the
      mature protein, was confirmed by protein sequencing
      This protein is involved in oral morphogenesis preceding binary
      fission, and in nuclear events during fertilization, such as
      formation of gametic pronuclei and zygote formation of gametic
      pronuclei. It also acts as a mitochondrial enzyme, citrate
      synthase.

GENETICS
#genetic_code SGC5
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; cytoskeleton; mitochondrion;
      oxo-acid-lyase

FEATURE
1-21 #domain transit peptide (mitochondrion) #status
      predicted #label TNP\
22-462 #product 14-nm filament protein/citrate synthase #status
      experimental #label MAT

SUMMARY
#length 462 #molecular-weight 52575 #checksum 9423
      38.3%; Score 1329; DB 2; Length 462;
Query Match Best Local Similarity 46.8%; Pred. No. 8.43e-229;
Matches 207; Conservative 95; Mismatches 124; Indels 16; Gaps 14;

Db 25 NLKK-VIAEIIPOKQAELEKVEKYGDKVVGQYTVKQVIGMGMKGLMSDLSRC-DPYQ 82
      : : : : | | | | : | | : : | | : | | | | | | : : | | | | | | |
QY 39 DLRLSEVOELIPEQDRLKRIKSDMKGSI-GNITVDMVLGGMRGTMGLLWK-PHYLDPDE 96
      ||| :||| : : | | | | | : |
Db 83 GIIFRGYTTIPLKREFLPADPKAAQDANPELPEGIFWLLMTGQLTHAQVDALKHEWQN 142
      ||| :||| : : | | | | | : |
QY 97 GIRFRGLSIPECCQKVLPA-AKPG-G-EPLPEGLLMLLTGTGKVPSEQVNSIVSGIAE 150
      ||| :||| : : | | | | | : |
Db 143 RGTVNQDCVNFILN-LPKDLHSMTLSMALLYLDKDSFKALYDEGKISKDYDEPYED 201
      : : : : | | | | : : : : | | : : : : | | | | | | | | | | | |
QY 151 SGIISLIIMYTTIDALPVTAPHTQPTFATGCMALQVQSEFQKAYEKG-IHKSQKYWEPTYED 209
      : : : : | | | | : : : : | | : : : : | | | | | | | | | | | |
Db 202 SMOLIAKIPVAAIIYRHKYRDSKLIDSDSKLDWAGNVAHMMGPEOHVVKPCI-RGYLSI 260
      : : : : | | | | : : : : | | : : : : | | | | | | | | | | | |

```



```

SUMMARY      #length 397  #molecular-weight 44830  #checksum 904
Query Match   7.6%; Score 263; DB 2; Length 397;
Best Local Similarity 30.1%; Pred. No. 4.28e-24;
Matches 95; Conservative 73; Mismatches 118; Indels 30; Gaps 22;

Db 57 LLIWKLPTQAEIEFEVEIRTHRIKYHIR-DMMKCPETGHPMDALQTSAAALGL--- 112
   ||: ||: ||: : : : | | | | | | | | | | | | | | | | | | | |
Qy 129 LLLTCKVPSKEQVNSIVSGIAESGIISLIIMYTTIDALPVTAHPTQFATGVMAIQVQSE 188
   ||: ||: ||: : : : | | | | | | | | | | | | | | | | | | | |

Db 113 FY-A-RRALDDPKY-IRA--AVVRLAKIPTMVA-AFH-MIREGNDPIQPNDKLDYASNF 165
   | | : : : . | | . : | | : | : | : | : | : | : | : | : |
Qy 189 FOKAYEKGIIHKSRYWEPTEDSMNLIAQVPLVAAVYVRYMYKNG-DTIPKDESLDYGANF 247
   || | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 166 LYMLTEKEPDPPFAAKVDFVCLTLHAEH-TMNASTFSARVTASTLTDYPYAVVASAVGTLAG 224
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 248 AHMLGFSSE-WHELLMRLVYTIHSDHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAG 306
   || | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 225 PLHGANEVL---NMLEIG-SV--ENVRPYVEKCLANKORIMFGHRYVYKVKDPRAI 277
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 307 PLHGLANQEVLLWIKSVVEECENISKEQLKDYVWKTLSNGKVVPFGHGVLRKTVPRYT 366
   || | | | | | | | | | | | | | | | | | | | | | | | | | |

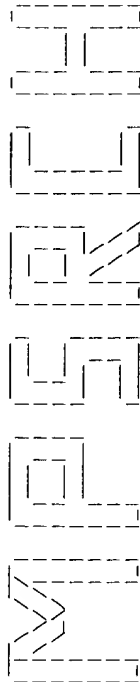
Db 278 ILQDLAEQLFAKMGHDEYIEIAVELEKVVVEYV-G-QK-GIYPNVDFYSG-LV-YRKLDI 332
   : : : | : | : : | : : | : : | : : | : : | : : | : : |
Qy 367 COREFA--M-KHLPEDPLFQLVSKLYEVEFLLEQLNLAKLKPWPVNDASHGVLLNYIGL-T 422
   : : : | : | : : | : : | : : | : : | : : | : : | : : |

Db 333 PADLFTPLFAIARVAG 348
   | | | | | | | | | |
Qy 423 EARYTVLFGVSRALG 438
   | | | | | | | | | |

```

Search completed: Fri Oct 22 16:31:22 1999
Job time : 158 secs.





 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Oct 22 16:31:39 1999; MasPar time 14.47 Seconds
 920.365 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-702-718-2
 Description: (1-471) from US08702718.pep
 Perfect Score: 3466
 Sequence: 1 MVFYRSVLSKLSRAVQO.....PLERPKSVTMEWLENOCKKA 471

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 i:swissprot

Statistics: Mean 51.825; Variance 92.291; scale 0.562

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2779	80.2	471	1	CISY_CITMA CITRATE SYNTHASE, MITO	0.00e+00
2	2305	66.5	472	1	CISY_ARATH CITRATE SYNTHASE, MITO	0.00e+00
3	1921	55.4	464	1	CISY_PIG CITRATE SYNTHASE, MITO	0.00e+00
4	1842	53.1	431	1	CISY_CHICK CITRATE SYNTHASE, MITO	0.00e+00
5	1722	49.7	468	1	CISY_CAEEL PROBABLE CITRATE SYNTH	0.00e+00
6	1716	49.5	482	1	CISY_SCHPO PROBABLE CITRATE SYNTH	0.00e+00
7	1663	48.0	479	1	CISY_YEAST CITRATE SYNTHASE, MITO	0.00e+00
8	1652	47.7	474	1	CISY_EMENI CITRATE SYNTHASE, MITO	0.00e+00
9	1642	47.4	475	1	CISY_CANTR CITRATE SYNTHASE, MITO	0.00e+00
10	1639	47.3	467	1	CISY_NEUCR CITRATE SYNTHASE, MITO	0.00e+00
11	1606	46.3	469	1	CISY_YEAST CITRATE SYNTHASE, MITO	0.00e+00
12	1575	45.4	460	1	CISY_TETTH CITRATE SYNTHASE, MITO	0.00e+00
13	1329	38.3	462	1	CISY_YEAST CITRATE SYNTHASE, MITO	8.03e-269
14	1168	33.7	486	1	CISY_YEAST CITRATE SYNTHASE 3 (EC	3.93e-231
15	316	9.1	437	1	CISY_CORGL CITRATE SYNTHASE (EC 4	2.97e-39
16	307	8.9	436	1	CISY_ACEAC CITRATE SYNTHASE (EC 4	2.08e-37
17	269	7.8	428	1	CISY_PSEAF CITRATE SYNTHASE (EC 4	9.64e-30
18	267	7.7	411	1	CISY_RICCA CITRATE SYNTHASE (EC 4	2.41e-29
19	266	7.7	411	1	CISY_RICJA CITRATE SYNTHASE (EC 4	3.80e-29
20	266	7.7	516	1	CISY_CUCMA CITRATE SYNTHASE, GLYO	3.80e-29
21	264	7.6	372	1	CISY_STRMU CITRATE SYNTHASE (EC 4	9.47e-29
22	263	7.6	397	1	CISY_SYNY3 CITRATE SYNTHASE (EC 4	1.49e-28
23	263	7.6	411	1	CISY_ASTPI CITRATE SYNTHASE (EC 4	1.49e-28

Query match 80.2%; Score 2779; DB 1, Length 471;

24	262	7.6	411	1	CISY_RICSL CITRATE SYNTHASE (EC 4	2.35e-28
25	262	7.6	411	1	CISY_RICPA CITRATE SYNTHASE (EC 4	2.35e-28
26	262	7.6	411	1	CISY_RICSI CITRATE SYNTHASE (EC 4	2.35e-28
27	262	7.6	412	1	CISY_RICBE CITRATE SYNTHASE (EC 4	2.35e-28
28	260	7.5	411	1	CISY_RICCN CITRATE SYNTHASE (EC 4	5.85e-28
29	259	7.5	411	1	CISY_RICGE CITRATE SYNTHASE (EC 4	9.21e-28
30	257	7.4	411	1	CISY_RICRH CITRATE SYNTHASE (EC 4	2.28e-27
31	257	7.4	411	1	CISY_RICMA CITRATE SYNTHASE (EC 4	2.28e-27
32	255	7.4	411	1	CISY_RICAF CITRATE SYNTHASE (EC 4	5.64e-27
33	255	7.4	411	1	CISY_RICAU CITRATE SYNTHASE (EC 4	5.64e-27
34	253	7.3	411	1	CISY_RICTY CITRATE SYNTHASE (EC 4	1.39e-26
35	254	7.3	423	1	CISY_ACIAN CITRATE SYNTHASE (EC 4	8.86e-27
36	254	7.3	430	1	CISY_COXBU CITRATE SYNTHASE (EC 4	8.86e-27
37	249	7.2	373	1	CISY_BAGCO CITRATE SYNTHASE (EC 4	8.44e-26
38	249	7.2	411	1	CISY_RICAK CITRATE SYNTHASE (EC 4	8.44e-26
39	251	7.2	431	1	CISY_MYCTU CITRATE SYNTHASE 1 (EC	3.43e-26
40	245	7.1	436	1	CISY_RICPR CITRATE SYNTHASE (EC 4	5.08e-25
41	242	7.0	372	1	CISW_BACSU CITRATE SYNTHASE III (1.94e-24
42	238	6.9	431	1	CISY_BARHE CITRATE SYNTHASE (EC 4	1.15e-23
43	230	6.6	376	1	CISY_PYRFU CITRATE SYNTHASE (EC 4	3.98e-22
44	226	6.5	427	1	CISY_ECOLI CITRATE SYNTHASE (EC 4	2.31e-21
45	227	6.5	429	1	CISY_RHTR CITRATE SYNTHASE, CHRO	1.49e-21

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	471 AA.
ID	CISY_CITMA			
AC	P49298;1996 (REL. 33, CREATED)			
DT	01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)			
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)			
DE	CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).			
GN	CIT			
OS	CITRUS MAXIMA (PUMMEL) (CITRUS GRANDIS).			
OC	EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			
OC	SAPINDALES; RUTACEAE; CITRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV. STAMSE SWEET 2240;			
RA	CANEL C.;			
RL	THESIS (1994), UNIVERSITY OF CALIFORNIA AT RIVERSIDE, U.S.A.			
CC	-!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +			
CC	OXALOACETATE.			
CC	-!- PATHWAY: TRICARBOXYLIC ACID CYCLE.			
CC	-!- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.			
CC	-!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE			
CC	METABOLISM.			
CC	-!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U19481; 6245676; ..			
DR	PROSITE; PS00480; CITRATE SYNTHASE; 1.			
DR	PFAM; PF00285; citrate_synt; 1.			
DR	HSP; P23007; 5CSC.			
KW	LVASE; TPICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.			
FT	TRANSIT 1 18 MITOCHONDRION (POTENTIAL).			
FT	CHAIN 19 471 CITRATE SYNTHASE.			
FT	ACT_SITE 307 307 BY SIMILARITY.			
FT	ACT_SITE 353 353 BY SIMILARITY.			
FT	ACT_SITE 408 408 BY SIMILARITY.			
SQ	SEQUENCE 471 AA; 52183 MW; A7E3512D CRC32;			

Best Local Similarity 78.4%; Pred. No. 0.00e+00;
Matches 371; Conservative 56; Mismatches 41; Indels 5; Gaps 5;

Db 1 MASLRATASRLRSPRAGQSSNLSNVPWQMOSSADLHSQL-KEMIPQEQELKKVK 59
QY 1 MVEYRSVSLSLKSRSAVOQSNVSVRWLQVOTSSGLDLSELVQELPEQDRLKKIK 60

Db 60 SDLGKAGLQNLITDVIIGMGRMTGLLWETSLDDPEGRFGLSIPCEQKLLPAAPDG 119
QY 61 SDM-KGSIGNITVDMVLGGMGRMTGLLWPKPHYLDPEGRFGLSIPCEQKVLPAAPGG 119

Db 120 EPLPEGLLLWLLTGKVPKSEQVGLSKELRDRAVDPDVV-YKAIDALPVSAPHPMTQFASG 178
QY 120 EPLPEGLLLWLLTGKVPKSEQVGLSKELRDRAVDPDVV-YKAIDALPVSAPHPMTQFASG 179

Db 179 VMAQVQSFQAEYKGIHKSWEPTSDSUNLIARVPVVAAYVQRIYKDGKIIPKDD 238
QY 180 VMAQVQSFQAEYKGIHKSWEPTSDSUNLIARVPVVAAYVQRIYKDGKIIPKDD 239

Db 239 SLDYGCNFSHMLGFDPKMLEL-MRLYVTIHSDEHGGNVSARTGHLVASALSDPYLSFLA 297
QY 240 SLDYGCNFSHMLGFDPKMLEL-MRLYVTIHSDEHGGNVSARTGHLVASALSDPYLSFLA 299

Db 298 ALNGLAGPLHGLAQEVLLWIKSVVDECGENVTTTQLKDYVWKTINSKVVYPFGHGVLR 357
QY 300 ALNGLAGPLHGLAQEVLLWIKSVVDECGENVTTTQLKDYVWKTINSKVVYPFGHGVLR 359

Db 358 KTDPRYTCOREFALKHLPDDPLFQVLSKLYEVPPIILKGLKVNKNPNVDHAGSVLLNH 417
QY 360 KTDPRYTCOREFALKHLPDDPLFQVLSKLYEVPPIILKGLKVNKNPNVDHAGSVLLNH 418

Db 418 FGLAEARYTVTLFGVSRSLGICSLIDRALGLPLERPKSVTLDWIEKNCKKA 470
QY 419 YGLTEARYTVTLFGVSRSLGICSLIDRALGLPLERPKSVTLDWIEKNCKKA 471

RESULT 2
ID C1SY ARATH STANDARD; PRT; 472 AA.
AC P20115;
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7)
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPALANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPAELES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91370823.
RA UNGER E.A., HAND J.M., CASHMORE A.R., VASCONCELOS A.C.;
RT "Isolation of a cDNA encoding mitochondrial citrate synthase from
Arabidopsis thaliana."
RL PLANT MOL. BIOL. 13:411-418(1989).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17528; E1188578;
PIR: JA0149; YKMDM.

DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
DR HSP: P23007; 5CSG.
KW LYSASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
KW MULTIGENE FAMILY.
FT TRANSIT 1 ? MITOCHONDRION.
FT CHAIN ? 472 CITRATE SYNTHASE.
FT ACT_SITE 309 309 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 407 407 BY SIMILARITY.
SQ SEQUENCE 472 AA; 52941 MW; 954AFA81 CRC32;

Query Match 66.5%; Score 2305; DB 1; Length 472;
Best Local Similarity 68.9%; Pred. No. 0.00e+00;
Matches 328; Conservative 60; Mismatches 76; Indels 12; Gaps 12;

Db 1 MVFFRSVSAFTRLSRVGOQSSLSNSVRWIMQSSSTDLQKSL-QELIPEKHDLKKL 59
QY 1 MVFYRSVSLSLKSRSA-VQSNVSNVSRWLQVOTSSGLDLSELVQELIPEQDRLKKI 59

Db 60 KSEHGKVLGNITVDMVIGMGRMTGLLWETSLDDPEVFALGDCRLPECEKALLPTAQS 119
QY 60 KSDM-KGSIGNITVDMVLGGMGRMTGLLWPKPHYLDPEGRFGLSIPCEQK-VLPAAPK 117

Db 120 GCLNHYRSEFVASLWKGTL-AKSKLKH-CRKTWNRAVSDYV-YNAIDALPSTAHPTQ 176
QY 118 GG-EPLPGLL-WULLTGKVPKSEQVNSIVSGIAESGIISLIIMYTTIDALPVTAHPTQ 175

Db 177 FASGYMALQVQSEFQKAYENGIIHKSKEWPTYEDCLNLIARVPVVAAYVYRMYKNGDSI 236
QY 176 FATGYMALQVQSEFQKAYEKGIHKSKEWPTYEDSMNLIARVPVVAAYVYRMYKNGDTI 235

Db 237 PSDKSLDYGANFSHMLGDDDLRLKEL-MRLTSPSTVWHEGNGNVSARTGHLVGSALSDPYL 295
QY 236 PKDESIDYGANFAHMLGFSSEMHELLMRLYVTIHSDEHGGNVSARTGHLVGSALSDPYL 295

Db 296 SFAAALNGLAGPLHGLAQEVLLWIKSVVDECGEDISKEOLKEYVWKTINSKVVYPGYGH 355
QY 296 SFAAALNGLAGPLHGLAQEVLLWIKSVVDECGENISKEOLKDYVWKTINSKVVYPFGH 355

Db 356 GVLRTNDRYVQCFEFALKHHPDPLFQ-CKLMLK-ASCLTELESEEPNPNVDHAGSVL 413
QY 355 GVLRTNDRYVQCFEFALKHHPDPLFQ-CKLMLK-ASCLTELESEEPNPNVDHAGSVL 415

Db 414 LNHVGLTEARYTVTLFGVSRSLGICSLIDRALGLPLERPKSVTLDWIEKNCKKA 469
QY 416 LNYGLTEARYTVTLFGVSRSLGICSLIDRALGLPLERPKSVTLDWIEKNCKKA 471

RESULT 3
ID C1SY PIG STANDARD; PRT; 464 AA.
AC P00889;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
OS CS.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTILA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89000665.
RA EVANS C.T., OWENS D.D., SUMEGI B., KISPAL G., SRERE P.A.;
RT "Isolation, nucleotide sequence, and expression of a cDNA encoding
pig citrate synthase."
RL BIOCHEMISTRY 27:4680-4686(1988).
CC [2]
RP SEQUENCE OF 28-464.
RX MEDLINE: 82231993.
RA BLOXHAM D.P., PARMELEE D.C., KUMAR S., WALSH K.A., TITANI K.;
RT "Complete amino acid sequence of porcine heart citrate synthase."
RL BIOCHEMISTRY 21:2028-2036(1982).

RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.7 AND 1.7 ANGSTROMS).
 RX MEDLINE: 83010291.
 RA REMINGTON S., WIEGAND G., HUBER R.;
 RI "Crystallographic refinement and atomic models of two different forms
 RL of citrate synthase at 2.7- and 1.7-A resolution.";
 RN J. MOL. BIOL. 158:111-152(1982).
 RP [4]
 RX MUTAGENESIS.
 RA MEDLINE: 91104711.
 RL ALTER G.M., CASAZZA J.P., ZHI W., NEMETH P., SRERE P.A., EVANS C.T.;
 RI "Mutation of essential catalytic residues in pig citrate synthase.";
 RL BIOCHEMISTRY 29:7557-7563(1990).
 CC -/- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -/- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -/- SUBUNIT: HOMODIMER.
 CC -/- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -/- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -/- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M21197; G164419; -.
 DR PIR: A29966; YKPG.
 DR PDB: 1CTS; 16-JUL-88.
 DR PDB: 2CTS; 09-OCT-88.
 DR PDB: 3CTS; 09-OCT-88.
 DR PDB: 4CTS; 16-JUL-88.
 DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 KW LYASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; 3D-STRUCTURE;
 KW TRANSIT PEPTIDE; METHYLATION.
 FT TRANSIT 1 27 MITOCHONDRION.
 FT CHAIN 28 464 CITRATE SYNTHASE.
 FT MOD_RES 395 395 METHYLATION (TRI-).
 FT ACT_SITE 301 347
 FT ACT_SITE 347 402
 FT ACT_SITE 402 402
 FT HELIX 33 55
 FT TURN 56 57
 FT TURN 59 64
 FT STRAND 65 69
 FT TURN 70 71
 FT STRAND 76 79
 FT STRAND 84 86
 FT TURN 87 89
 FT STRAND 90 92
 FT STRAND 97 97
 FT HELIX 98 104
 FT STRAND 107 107
 FT STRAND 114 114
 FT TURN 116 116
 FT TURN 117 125
 FT HELIX 131 141
 FT TURN 142 143
 FT HELIX 149 157
 FT TURN 160 161
 FT TURN 164 165
 FT TURN 166 175
 FT TURN 176 177
 FT HELIX 181 187
 FT TURN 188 188
 FT TURN 192 193
 FT HELIX 194 221
 FT TURN 223 224

FT TURN 232 233
 FT TURN 236 237
 FT HELIX 238 244
 FT TURN 245 245
 FT TURN 249 250
 FT HELIX 251 261
 FT TURN 262 262
 FT TURN 266 268
 FT HELIX 270 280
 FT TURN 281 282
 FT HELIX 285 296
 FT TURN 297 297
 FT TURN 299 309
 FT HELIX 310 318
 FT TURN 319 319
 FT TURN 325 328
 FT HELIX 329 337
 FT TURN 338 339
 FT STRAND 345 345
 FT HELIX 355 364
 FT TURN 365 367
 FT TURN 369 370
 FT HELIX 372 375
 FT TURN 376 379
 FT TURN 381 383
 FT HELIX 417 441
 FT TURN 442 443
 FT STRAND 451 452
 FT HELIX 454 461
 FT TURN 462 462
 SQ SEQUENCE 464 AA; 51629 MW; 3BC38D25 CRC32;
 Query Match 55.4%; Score 1921; DB 1; Length 464;
 Best Local Similarity 58.8%; Pred. No. 0.00e+00;
 Matches 254; Conservative 86; Mismatches 86; Indels 6; Gaps 6;
 Db 27 SASSTNLK-DILADLPKEQARIKTFRQOHGNTVVGITVDMYGGMRGKGLVYETSVL 85
 Qy 34 TSSGLDLRSELVQELIPQODRLKKIKSDMKGSI-GNITVDMVLGMRGHTGLLKKPHYL 92
 Db 86 DPDEGIRFGYSIPECQKMLPKAKGGEPLPEGLFWLLVTGQIPTERQVSWLSKEWAKRA 145
 Qy 93 DPDEGIRFGLSIPECQKVLPAKPGGEPLPEGLLWLLLTGKVPKSEQVNSIVSGIAESG 152
 Db 146 ALPSHVV-TMLDNFPTNLHPMSQLSAAITALNSESNTAFARAYAEIGIHRTKYWEIYECMD 204
 Qy 153 IISLIIMYTTIDALPVTAHPTQFATGMALQVQSEFQKAYEKGIGHKSKYWEPTYESMN 212
 Db 205 LIAKLPVAAKIYRNLYREGSSIGAIKSDKLDWNSHNTNMLGYTDAQFTEL-MRLYLTIHS 263
 Qy 213 LIAQVPLVAAVYRRMYKNGDTI-PKDESLDYGANFAHMLGFSSEMHELLMRLYVTIHS 271
 Db 264 DHEGNNVSAHTSHLVGSALSADPYLSFRAAMNGLAGPLHGLANQEVLYWLTOLOKEVQDV 323
 Qy 272 DHEGNNVSAHTGHLVASALSADPYLSFRAALNGLAGPLHGLANQEVLLW-KSVVEECGNI 331
 Db 324 SDEKLRYIWNLTNSGRVPGYGHAVLRKTDPRYTQREFALKHLPDPMFKLVAQLYKI 383
 Qy 332 SKEQLKDYVWKTLNSGKVVPGFGHVLKRTVRYTCQREFAMKHLPEPLFQLYSKLYEV 391
 Db 384 VPNVLLEOGKAKNPVNDVDAHSGVLLQYVGMTEYNNYTVLVFGVSRALGVLAQLIWSRALG 443
 Qy 392 FLLFLQNLAKLK-PWPNVDVDAHSGVLLNYGLTEARYTVLVFGVSRALGICSLQSLWDRALG 450
 Db 444 FPLRPKSMSTD 455

```

QY 451 LPLRPSVTIME 462
RESULT 4
ID C15Y_CHK1K STANDARD; PRT: 433 AA.
AC P23007;
DI 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL (EC 4.1.3.7).
GN CS.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE: 90248434.
RA KARPUSAS M., BRANCHAUD B., PEMINGTON S.J.;
RT "Proposed mechanism for the condensation reaction of citrate
RT synthase: 1.9-A structure of the ternary complex with oxaloacetate
RT and carboxymethyl coenzyme A.";
RL BIOCHEMISTRY 29:2213-2219(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF OPEN CONFORMATION.
RX MEDLINE: 91255228.
RA LIAO D.-I., KARPUSAS M., REMINGTON S.J.;
RT "Crystal structure of an open conformation of citrate synthase from
RT chicken heart at 2.8-A resolution.";
RL BIOCHEMISTRY 30:6031-6036(1991).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -!- CAUTION: THIS IS AN X-RAY DETERMINED SEQUENCE WHICH WAS
CC ESTABLISHED USING THE SEQUENCE OF PIG CITRATE SYNTHASE AND
CC MODIFYING IT BASED ON THE OBSERVED ELECTRON DENSITY.
DR PDB: 1CSC; 15-APR-91.
DR PDB: 2CSC; 15-APR-91.
DR PDB: 3CSC; 15-APR-91.
DR PDB: 4CSC; 15-APR-91.
DR PDB: 5CSC; 15-APR-91.
DR PDB: 6CSC; 24-DEC-97.
DR PDB: 5CTS; 15-OCT-90.
DR PDB: 6CTS; 15-JAN-93.
DR PDB: 1CSH; 30-APR-94.
DR PDB: 1CSI; 30-APR-94.
DR PDB: 1CSR; 15-OCT-95.
DR PDB: 1CSS; 15-OCT-95.
DR PDB: 1AL6; 24-DEC-97.
DR PDB: 1AMZ; 24-DEC-97.
DR PROSITE: P500480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; 3D-STRUCTURE.
FT ACT_SITE 274 274
FT ACT_SITE 320 320
FT ACT_SITE 375 375
FT HELIX 5 28
FT TURN 29 30
FT HELIX 38 42
FT TURN 43 47
FT STRAND 56 59
FT TURN 60 62
FT STRAND 63 66
FT TURN 67 68
FT STRAND 69 70
FT HELIX 71 77
FT STRAND 80 80

```

```

FT TURN 82 83
FT STRAND 87 87
FT HELIX 89 98
FT HELIX 104 117
FT HELIX 122 130
FT TURN 133 134
FT HELIX 137 151
FT HELIX 153 160
FT TURN 161 161
FT HELIX 164 194
FT TURN 196 197
FT TURN 205 206
FT HELIX 209 217
FT TURN 218 218
FT HELIX 222 234
FT TURN 235 235
FT HELIX 243 253
FT TURN 254 255
FT HELIX 258 269
FT TURN 270 270
FT TURN 272 276
FT HELIX 277 291
FT TURN 292 293
FT HELIX 298 310
FT TURN 311 312
FT TURN 316 317
FT STRAND 318 318
FT HELIX 328 340
FT TURN 342 343
FT HELIX 345 364
FT TURN 365 365
FT STRAND 372 372
FT TURN 374 374
FT HELIX 375 384
FT TURN 385 386
FT HELIX 390 414
FT TURN 415 416
FT HELIX 427 433
SQ SEQUENCE 433 AA; 47373 MW; 6E24FE58 CRC32;

```

Query Match 53.18; Score 1842; DB 1; Length 433;

Rest Local Similarity 57.18; Pred. No. 0.00e+00;

Matches 246; Conservative 85; Mismatches 94; Indels 6; Gaps 6.

```

Db 1 ASSTNLK-DYLAALIPKEQARIKTFROHGGTALGQITVDMSYGGMRGKGLVYETSVD 59
QY 35 SSGDLRLSELVQELIPEQDRLKKIKSDMKGSG-IGNITVDVVLGGRCMTGLLWKPHYLD 93
Db 60 PDEGIRFGFSIPECQKLLPKGXGGEPLPEGLFWLLVTGQITPGQVSWLSKEWAKRAA 119
QY 94 PDEGIRFGFSIPECQKLVPAKPGGEPLPEGLLWLLTGKVPKSGQVNSIVSGIAESGI 153
Db 120 LPSHVW-TMLDNEPTNLHPMSOLSAAITALNSESNFARAYAEGLRTKYWEMVYESAMD 178
QY 154 ISLIIMTTIDALPVTAPHTQFATGVMAIQVSEPOKAYEKGIHKSKEWYPTYESMNL 213
Db 179 IAKLPCVAAKIYRNLYRAGSSIGAKSDKLDWSHNTNMLGYTDAQFTTEL-MRLYLTIHSD 237
QY 214 IAQVPLVAAVYVRMYKNGDTI-PKDESLOYGANFAHMLGFSSEMHELLMRLYVYTHSD 272
Db 238 HEGGNVSAHTSHLVGSALSDDPYLSFAAAMNGLAGPLHGLANQEVGLWLAOLQKAXXAGA 297
QY 273 HEGGNVSAHTSHLVGSALSDDPYLSFAAALNGLAGPLHGLANQEVGLWIKSVVEECGENIS 332
Db 298 DASLRDVIWNTLNSGVPVPGYGHAVLRKTDPRYTCOPEFALKHLPDPMKFLVAOLYKIV 357
QY 333 KEQLKDYVWKTLSGKVPVPGFEGHGLRKTVPRYTCOREFAKKHLPEDPLFQLVSKLYEYF 392
Db 358 PNVLLQEGAAANPWPNDVDAHSGVLLQYTGTEMNYYTVLFVGSKALGVLAQLIWSRALGF 417
QY 393 L-LFLQNLAKLKPWPNDVDAHSGVLLNYGLTEARYTYTVLFGVSRALGICSLIDRALGL 451
Db 418 PLRPKSMSTD 428

```

```

|||||: :
452 PLERPKSVTME 462

RESULT 5
ID C1SY_CABEL STANDARD: PRT: 468 AA.
AC P34575;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN T2063.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BERKS M., SMITH A.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z30423; G458482; -.
DR PIR; S42370; S42370.
DR WORMPEP; T2063.2; CE00513.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; SCSC.
KW HYPOTHETICAL PROTEIN; LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION;
KW TRANSIT PEPTIDE.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 468 CITRATE SYNTHASE.
FT ACT_SITE 303 303 BY SIMILARITY.
FT ACT_SITE 349 349 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
SQ SEQUENCE 468 AA; 51540 MW; ADA3A630 CRC32;

Query Match 49.7%; Score 1722; DB 1; Length 468;
Best Local Similarity 54.0%; Pred. No. 0.00e+00;
Matches 236; Conservative 84; Mismatches 109; Indels 8; Gaps 8;

Db 25 LSTSAEGSTNLKVLKSK-IPAHNAKVKSPRTEHGSTVQONVNDIMYGGMRSMKGMVTE 83
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 30 LQVOTSSGLDLSELVQELIPEQQDLRLKIKSDMKSGIG-NITVDMVLGMRGMTGLLWK 88
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 84 TSVLDPEEGIRFRGYSIPKCKLLPKAKGGEPLPEALWLLCTGDVPSQAQTAATKEW 143
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 89 PHYLDPEGIRFRGLSPKQKVLPAKPGGEPLPEGLLWLLITGKVPKQVNSIVSGI 148
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 144 NARADLPTHVV-RMLDNFPNLPMAQIFAAIALNNESKFAAGYARGVAKASYWEVAYE 202
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 149 AESGIISLIWYTTIDALPVTAHPTQFATGVWALQVQSEFQAYEKGIHKSXYWEPTYE 208
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 203 DSMDLAKLPVAAITIRNLYRDSGASVSDIPKQWSANFSSMLGDDPLFAEL-WRLYL 261
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 209 DSMNLIAQVPLVAAYVRRYKNGDTIP-KDESIDYGANFAHMLGFSSEMHELLMRLYV 267
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 262 VIHSDHEGNGVSAHTSLVGSALSDPYLSFAAMAGLAGPLHGLANCEVLLWIKSVVEEC 321
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |

```

```

|||||: : : : :
Qy 268 TIHSDHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANCEVLLWIKSVVEEC 327
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 322 GFNYTEEQKKEWVKHLKSGQVPGYGHAVLRKTDPRYECOREFALKHLPNDLFLKLVST 381
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 328 GENISKEQLKDYWKTLNKGKVPFGHGVLRKTVPRYTICQREFAMKHLPEPLFQLYSK 387
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 382 LYKITPGILLFQGRKAKNPVNDVSHSGVLLQYFGMTSEFYTLVFGVSRLG-CLSOLIW 440
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 398 LYEVFLFLQNLAKLK-PWPNDVDAHSGVLLNYGLTEARYTVLFGVSRLGIC-SOLIW 445
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 441 ARGMGPLERPKSHSTD 457
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 446 DRALGLERPKSVTME 462
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |

RESULT 6
ID C1SY_SCHPO STANDARD: PRT: 482 AA.
AC Q10306;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROBABLE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN SPAC503.03.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69731; E223710; -.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; SCSC.
KW HYPOTHETICAL PROTEIN; LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION;
KW TRANSIT PEPTIDE.
FT TRANSIT 1 44 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 45 482 CITRATE SYNTHASE.
FT ACT_SITE 319 319 BY SIMILARITY.
FT ACT_SITE 365 365 BY SIMILARITY.
FT ACT_SITE 420 420 BY SIMILARITY.
SQ SEQUENCE 482 AA; 54030 MW; 18145A7D CRC32;

Query Match 49.5%; Score 1716; DB 1; Length 482;
Best Local Similarity 55.8%; Pred. No. 0.00e+00;
Matches 237; Conservative 83; Mismatches 101; Indels 4; Gaps 4;

Db 51 LKADRLA-ELIPEKQAEIKKRAEHGQDVIGEVINQWVGARGVRSLLNEGSVLDPNEG 109
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 40 LRSELVQELIPQOORLKKIKSDM-KGSIGNITVDMVLGMRGMTGLLWIKPHYLDPEGI 98
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 110 RFRGTYIPECCQLPSPNGKPLPESLFWLLVTGEITLTSQVQALSADWAARSOLPKFV 169
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |
Qy 99 RERGLSIPCCQKVLPAKPGGEPLPEGLLWLLITGKVPKQVNSIVSGIAFGIISLI 158
| . . . | . | . . . | . . . | . . . | . . . | . . . | . . . |

```

Db	170	E-ELIDRCPTLHPMAQFSVLAVTALBHDSSAFKAKAYERGMMKHHDYWKYEVYEDCDMLIAKT	228
QY	159	MYTIDLPVTAHPMTQFATGVNALGVQSEFOKAEKGJHKSKYWEPTYEDYSMNLIQAOP	218
Db	229	PIAGRTYRNLVYRGCVVAPVQMDKXHSYNFANVLGFANNEEFVELMRLYLTIHADHEG	288
QY	219	LVAAYVYRMYKNGDITIPDESIDYCANFAHMLGFSSEMHELLMRLYVTHSDHEG	278
Db	289	SAHTGHLVGSALSPPFLSMAASLINGLAGPLHGLANOEVNLFILTKKEIGDDLSSETKS	348
QY	279	SAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANOEVLLWTKSVVEGGENISKEQLKD	338
Db	349	YLAKLINSRGVVPYGYGHAVLKRTPPYTAQRFALHEHLKPDPMFOLVSPYIYVPGVLTE	408
QY	339	YVAKTLNSRGVVPYEGHGVLKRITVPRYTTCOREFAMKHLPEDDLFLQVLSKLYEVFLFLQ	398
Db	409	HGKTKPNYPNVDSHSGVLLQYCYCLKEQSEYTVLFGVSRITLGVASQLINDRALGLPIERP	468
QY	399	LAKLK-PWPNVDHAHSGVLLNYGLTEARVYTVLFGVSRALGTCSQLINDRALGLPIERP	457
Db	469	SPSTE 473	
QY	458	SVTME 462	

RESULT

```

ID      CDS_YEAST      STANDARD:      PRG:      479  AA.
AC      P00890;
DT      21-JUL-1986 (REL. 01, CREATED)
DT      01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT      01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DI      CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
DE      CITL OP LY56 OR GLU3 OP YN5001C OR W2019.
GN      SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OS      SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC      EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC      SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
RN      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC      STRAIN-D273-10B;
RA      LINDNER P., PLUECKTHUN A.;
RL      SUBMITTED (JUL-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN      SEQUENCE FROM N.A.
RC      STRAIN-D273-10B;
RX      MEDLINE: 85003587.
RA      SUISSA M., SUDA K., SCHATZ G.;
RT      "Isolation of the nuclear yeast genes for citrate synthase and
RT      fifteen other mitochondrial proteins by a new screening method ";
RL      EMBO J 3:1773-1781(1984).
[3]
RN      SEQUENCE FROM N.A.
RC      STRAIN-S288C / FY1679;
RX      MEDLINE: 95208356.
RA      VERHASSEL P, AERT R., VOET M, VOLCKAERT G.;
RT      "Twelve open reading frames revealed in the 23.6 kb segment flanking
RT      the centromere on the Saccharomyces cerevisiae chromosome XIV right
RT      arm.";
RL      YEAST 10:1355-1361(1994)
[4]
RN      SEQUENCE FROM N A
RC      STRAIN-GRF88;
RX      MEDLINE: 95028151.
RA      LALO D., SPETTLER S, MAPIOTTE S., GENOPEAU E., THURIAUX P.;
RT      "Organization of the centromeric region of chromosome XIV in
RT      Saccharomyces cerevisiae.";
RL      YEAST 10:523-533(1994).
CC      -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC      OXALOACETATE.
CC      -!- PATHWAY: TRICARBOXYLIC ACID CYCLE
CC      -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC      -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC      METABOLISM.

```

```

15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN CITA.
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; EUPICELLA
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FGSC 4;
RX MEDLINE; 97306446;
RA PARK B.W., HAN K.H., LEE C.Y., LEE C.H., MAENG P.J.;
RT "Cloning and characterization of the cita gene encoding the
RL mitochondrial citrate synthase of Aspergillus nidulans.";
MO. CELLS 7:290-295(1997).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U89675; G2138332;
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; 5CSC.
KW LYSASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 35 MITOCHONDRION (POTENTIAL).
FT CHAIN 36 474 CITRATE SYNTHASE.
FT ACT_SITE 310 310 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
SQ SEQUENCE 474 AA; 5223 MW; D97E1E91 CRC32;

Query Match 47.7%; Score 1652; DB 1; Length 474;
Best Local Similarity 54.1%; Pred. No. 0.00e+00;
Matches 246; Conservative 85; Mismatches 116; Indels 8; Gaps 7;

Db 15 STLAAPVQSVAFNGLR--CYSTGKTKSLKETFADKL-PGELEKVKKLKEHGNKVIGE 71
QY 11 SKLSRAVQSVNSVRVWLVQVTSGLDLRSELVQELIPEQDRLKIKSDMKGS-IGN 69

Db 72 LTLDQAYGGARGVKLVWEGSVLDSEEGIRFRLTIPEQCKLLPKAPGGEPLPEGLFWL 131
QY 70 ITVDMVLGMRGNTGLLWPKHYLDLDEGIRFRLTIPEQCKLVPAKPGGEPPEGLLWL 129

Db 132 LTGVEPSEQVQVRLSAEWAASDLPKFTL-ELIDRVSTLHPMAQFSLAVTALHESAF 190
QY 130 LTRGKVPSEQVNSVSGAEGSIILTIIMYTIDALPVTAHPMTQFATGVNALQVSEF 189

Db 191 AKATAGINKDKWNYTFEDSDMLIAKLPTIAAKIYRNVFKDGKVPKIQDKDYSYNLAN 250
QY 190 QKAYENGHKSKEWTEYEDSMNLIAQVPLVAAYVYRMYKNGDIIPKDESLDYGANFAH 249

Db 251 OLGFAONKDFEVL-MRLYITIHSDHEGNNVSAHTHLVGSALSSPMLSLAAGNLGLAPL 309
QY 250 MLGFSSS-EMHLEMLRYLTIHSDHEGNNVSAHTHLVGSALSDPYLSFAAALNGLAPL 308

Db 310 HGLANQEVNLWLTETEMKKVGVNDLSQSIKDYLNWSTLNAGRVVPGYGHVLRKTDPRYSQ 369
QY 309 HGLANQEVNLWLVKSVVEEGENISKQADYWKTLNSGKVPVGGHGVLRKIVPRYTCQ 368

Db 370 REFALRKLDDPDMFKLVQSVYKIAPGVLTGKTKNPVNDVDAHSGVLLIQQYGLTERNY 429
QY 369 REFAMKHLPEDLPLQVLSKLYEVFLLQLNLAKLK-PWPNVDHSGVLLNYYGLTEARY 427

```

```

430 TVLFGVSRALGVLPOLLIDRAFCAPIERPKSFSTE 464
QY 428 TVLFGVSRALGICQLIWDRLGLPLERPKSVTWE 462

RESULT 9
ID CYSY ASPNG STANDARD; PRT; 475 AA.
AC P51044;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7)
GN CIT-1.
OS ASPERGILLUS NIGER.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WU-2232L;
RA OSHIDA Y., MIYAKE K., KANAYAMA S., KIRIMURA K., USAMI S.;
RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63376; G927641;
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; 5CSC.
KW LYSASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 475 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 310 310 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
SQ SEQUENCE 475 AA; 52153 MW; 48D237BC CRC32;

Query Match 47.4%; Score 1642; DB 1; Length 475;
Best Local Similarity 54.9%; Pred. No. 0.00e+00;
Matches 237; Conservative 82; Mismatches 107; Indels 6; Gaps 6.

Db 36 TGRAKSLKETFAEKL-PAELEKVKKLKEHGSVIGEVLDQAYGGARGVKLVWEGSVL 94
QY 34 TSSGLDLRSELVQELIPEQDRLKIKSDMKGS-IGNITVDMVLGMRGNTGLLWPKHYL 92

Db 95 DSEEGIRFRLTIPEQCKLLPKA-PGQEPLEGLFWLLTGTIPEQVQVPLSAEWAAP 153
QY 93 DPDEGIRFRLTIPEQCKLVPAKPGG-EPLPEGLLWLLTGTGKVPSEQVNSIVSGIAES 151

Db 154 SLPKFTIE-ELIDRCPTLHPMSQFSLAVTALHESAFKAYAKGINKKDYNYTFEDSM 212
QY 152 GIISLIIMYTIDALPVTAHPMTQFATGVNALQVSEFQKAYEKGIHKSKEWTEYEDSM 211

Db 213 DLIAKLPTIAAKIYRNVFKDGKVPKIQDKDYSYNLANQGLYGDNDNFVLMRLYLIHS 272
QY 212 NLIAQVPLVAAYVYRMYKNGDIIPKDESJDYGANFAHMLGFSSEMHEMLMLRYLTIHS 271

Db 273 DHEGNNVSAHTHLVGSALSSPMLSLAAGNLGLAPLHGLANQEVNLWLTMKKAAIGNDL 332
QY 272 DHEGNNVSAHTHLVGSALSDPYLSFAAALNGLAPLHGLANQEVNLWLVKSVVEEGENI 331

```

Db 333 SDEAIKNYLWSTLNAGOVVPGYGHAVLRKTDPRVVSQREFALRKLDDPMFKLVSYQYKI 392
 QY 332 SKEQKDYVYKMTLNSGVVPGFVGHVLRKTVPRYTCORFAMKMLPDLPLFQVLSKLYEV 391
 Db 393 APCVTEHGKTKPNYPNDVHSGVLLQYGLTEANYTVLFGVSRALGVLPQLIIDRALG 452
 QY 392 FLLFQNLAKLK-PNPVNDVHSGVLLNYGLTEARYTVLFGVSRALGICLSQLIIDRALG 450
 Db 453 APIERPXSYSSTE 464
 QY 451 LPLERPXSYSSTE 462

RESULT 10
 ID C1SY_NEUCR STANDARD: PPT: 467 AA.
 AC P79024;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT.
 OS CANDIDA TROPICALIS (YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC CANDIDACEAE; CANDIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PK233;
 RA UEDA M., SANUKI S., KAWACHI H., SHIMIZU K., ATOMI H., TANAKA A :
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB001565; GI871439; -
 DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 DR HSSP: P23007; 5CSC.
 KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
 FT TRANSIT 1 ?
 FT CHAIN ? 467
 FT ACT_SITE 301 301 CITRATE SYNTHASE.
 FT ACT_SITE 347 347 BY SIMILARITY.
 FT ACT_SITE 402 402 BY SIMILARITY.
 FT ACT_SITE 402 402 BY SIMILARITY.
 SQ SEQUENCE 467 AA; 52004 MW; FEE342A9 CRC32;
 Query Match 47.3%; Score 1639; DB 1; Length 467;
 Best Local Similarity 54.0%; Pred. No. 0.00e+00;
 Matches 245; Conservative 80; Mismatches 117; Indels 12; Gaps 11;

Db 9 RSNVAKSTLKNVSR--TYATAEP-TLKQRL-BEILPAKAEVQKLDKDYKGVIGEVLL 64
 QY 14 RSRVQDSNVSNVSWVQVSSGLDRLSELVQELIPEQODRLKKIKSDM-KGSIGNITV 72
 Db 65 EOAYGCMGKICLVWEGSVLPDIEGIRPGRITPIDIQKELPKAPGGEPLPEALFWLLLT 124
 QY 73 DMVLGCMRGMTGLLWPKPHYLDPDEGIRPGRISLPEQOKVLPAAKPGGEPLPEGLLLMLLT 132
 Db 125 GEVPTAEQTRALSEFAARSALPKHVE-ELIDRSPSHLHPMAQFSIAVTALESSEOFAKA 183
 QY 133 GKVPSEQVNSVSGIAESGIIISLIIMYTTIDALPVTAPHTQFATGVMAQVQSEFOKA 192

Db 184 YAGVHKSEYWKYTYEDSIELLAKLPITIAAKIYRNVPDHQKLPQAIDSKLDYCANIASLL 243
 QY 193 YEGTHSKSYWPEYTYEDSMNLIAQVPLVAAVYVRRMYKNGD-TIPKDESIDYCANFAHML 251
 Db 244 GFGENKFELE-MRLYLTIIHSDHEGNGVSAHTHLVGSALSSPFLSLAAGLGLAGPLHG 302
 QY 252 GFSSS-EMHELMLRLYTIHSDHEGNGVSAHTHLVGSALSDPYLSFAALNGLAGPLHG 310
 Db 303 RANQEVLEWLFKLELNGDYSKAEIKYLDWTLNAGRVGPGYGHAVLRKTDPRYTAQRE 362
 QY 311 LANQEVLLWIKSVVEGGENISKQALDYVAKTLNSGVVPGFVGHVLRKTVPRYTCORE 370
 Db 363 FALKHMPDYELFKLVSNIEYVAPGVFDQHGMTK-NWPNVNGSHSGVLLQYVGLTESFYT 421
 QY 371 FAMKHPEDPLFQVLSKLYEVEL-LFLQN-LAKLKTWPNVDVHSGVLLNYGLTEARYYT 428
 Db 422 VLFVGSRAFGVLPOLIIDRGLMGPIERPXSYSSTE 455
 QY 429 VLFVGSRALGICLSQLIIDRALGPLERPXSYSSTE 462

RESULT 11
 ID C1SY_NEUCR STANDARD: PPT: 469 AA.
 AC P34085;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7)
 GN CIT-1.
 OS NEUROSPORA CRASSA.
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
 OC SORDARIALES; SORDARIACEAE; NEUROSPORA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-74A;
 RX MEDLINE: 94104594.
 RA FEREA T., CONTRERAS E.T., QUNG T., BOWMAN E.J., BOWMAN B.J.:
 RT "Characterization of the cit-1 gene from Neurospora crassa encoding
 RT the mitochondrial form of citrate synthase."
 RL MOL. GEN. GENET. 242:105-110(1994).
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -!- DEVELOPMENTAL STAGE: ABUNDANT AFTER 6-12 HRS OF GROWTH. IT IS
 CC NOT SIGNIFICANTLY EXPRESSED AFTER 24 HRS, WHICH IS SEVERAL HRS
 CC AFTER ENTERING THE STATIONARY PHASE OF GROWTH.
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M84187; GI68775; -
 DR PIR: S41563; S41563.
 DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 DR HSSP: P23007; 5CSC.
 KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
 KW MULTIGENE FAMILY.
 FT TRANSIT 1 33
 FT CHAIN 34 469 MITOCHONDRION (POTENTIAL).
 FT ACT_SITE 352 352 CITRATE SYNTHASE.
 FT ACT_SITE 352 352 BY SIMILARITY.
 SQ SEQUENCE 469 AA; 52002 MW; E6281B97 CRC32;
 Query Match 46.3%; Score 1606; DB 1; Length 469;


```

QY    330 NISKEQLKDVWTLNSGKVVGFGHGVLRNKTVPRTCTOREFAMKKHLPEDPLEQLVSKLY 389
Db    381 DVIPKKLLTYKKIANPNVDCHSGVLVSLGLTEYYQTTVTFVAVSRALG-CMANLIWSR 439
      :|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY    390 EVFLFLQNLAKL-KWPNPVDAHSGLNYGLTARYYTIVLVFGVSPLAGIC-SQLINDP 447
Db    440 AFGLPPIERPGSADLKWFHDXYR 461
      |..||...|| | . . .
QY    448 ALGLPLERPKSVTMEWLENCK 459

RESULT 14
ID CDSX-YEAST STANDARD; PRT; 486 AA.
AC P43635;
DI 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE 3 (EC 4.1.3.7).
GN CIT3 OR YPR001W OR YP9723.O1.
OS SACCHAROMYCES CERVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
PP SEQUENCE FROM N.A.
RC STRAIN-CW04;
RA JTA Y.K., BECAM A.-M., SLONIMSKI P.P., HERBERT C.J.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB372;
RA PEARDSON D.C., BOWMAN S., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RN SUBMITTED (APR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
[3]
PP SEQUENCE FROM N.A.
RA WANG Y., AHMED A., BUSSEY H., FORTIN N., PRIENSEN J.D., HALL J.,
RA STORMS R.K., VO D.H., WINNETT E.;
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -|- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-CoA + H(2)O + OXALOACETATE.
CC -|- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -|- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE METABOLISM.
CC -|- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by individuals or entities requires a license agreement (See http://www.isb-sib.ch/annotation or send an email to license@isb-sib.ch).
-----
DR EMBL; X88846; G895889; -
DR EMBL; Z48951; G763000; -
DR EMBL; Z71255; E236811; -
DR EMBL; U31900; G939735; -.
DR SGD; L000285; CIT3.
DR PROSITE; PS00342; MICROBODIES_CTER; 1.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; SSCS.
KW LYASE, TRICARBOXYLIC ACID CYCLE, MULTIGENE FAMILY.
FT ACT_SITE 315 315 BY SIMILARITY.
FT ACT_SITE 361 361 BY SIMILARITY.
FT ACT_SITE 419 419 BY SIMILARITY.
FT SITE 484 486 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 486 AA; 53811 MW; 806F987D CRC32;

Query Match 33.78; Score 1168; DB 1; Length 486;
Best Local Similarity 45.18; Pred. No. 3,93e+231;
Matches 207; Conservative 98; Mismatches 125; Indels 29; Gaps
Db 21 IIKSSALTLEAL-ENVIPKRDRVAKKLKACYGTGFITTSVLGGMGNGSMFWOGT 79

```

Search completed: Fri Oct 22 16:32:36 1999
Job time : 57 secs.



(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Sat Oct 23 15:30:25 1999; MasPar time 3195.79 Seconds
 Tabular output not generated. 1386.485 Million cell updates/sec

Title: >US-08-702-718-1
 Description: (1-1891) from US08702718.seq
 Perfect Score: 1891
 N.A. Sequence: 1 TTTTTCGTCCTCCATCAGCCTA.....TTATTGTGATATTATGAA 1891
 Comp: AAAAAGCAGGAGTAGTCGGAT.....AATAAAGTACTATATATCT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb1-est58
 1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
 6:em_est9 7:em_gss1
 Database: genbank-est11
 8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
 13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
 17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
 21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
 25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
 29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
 34:gb_est35 35:gb_est36 36:gb_est37 37:gb_gss1 38:gb_gss2
 39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.716; Variance 1.945; scale 5.023

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	241	12.7	471	20	AA896953 L30-386T3 Ice plant La	0.00e+00
2	200	10.6	523	17	T44185 7448 Lambda-PRL2 Arabi	0.00e+00
3	189	10.0	478	31	R90544 16899 Lambda-PRL2 Arab	0.00e+00
4	179	9.5	620	26	AU002699 A0002699 Bombyx mori p	0.00e+00
5	172	9.1	417	24	C96653 C96653 Rice callus Ory	0.00e+00
6	173	9.1	676	23	A1176862 EST220465 Normalized R	0.00e+00
7	153	8.1	657	15	AA606966 vm94h09.r1 Knowles Sol	3.25e-287
8	150	7.9	307	10	AA231705 CDO534.R cDNA from oat	5.79e-280
9	150	7.9	527	31	R90561 16916 Lambda-PRL2 Arab	5.79e-280
10	144	7.6	501	11	AA313713 EST185581 Colon carc.in	1.68e-265

	11	136	7.2	522	10	AA237580	mxild01.r1 Soares mous	2.69e-246
	12	125	6.6	213	8	217455	ARTS0029 AC16H Arabido	4.57e-220
	13	120	6.3	466	36	AA112941	zn59b06.r1 Stratagene	3.23e-208
	14	115	6.1	582	17	AA720242	33417 Lambda-FRL2 Arab	2.02e-196
	15	116	6.1	585	22	AI054878	cosu0002G14 Cotton Bol	8.91e-199
	16	113	6.0	473	14	AA498291	vh38g11.r1 Barstead mo	1.02e-191
	17	109	5.8	540	24	AI238915	GH15054.5prime GH Dros	2.42e-182
	18	105	5.6	442	33	W01297	yz94h09.r1 Soares mela	5.25e-173
	19	106	5.6	444	10	AA245100	mx08g10.r1 Soares mous	2.46e-175
c	20	106	5.6	489	23	AI168709	ox56a01.s1 Soares_NHMH	2.46e-175
	21	104	5.5	418	34	W40867	mc66e02.r1 Soares mous	1.12e-170
c	22	104	5.5	494	18	AA830955	oc58d01.s1 NCI_CGAP GC	1.12e-170
	23	104	5.5	503	9	AA177425	mt16h01.r1 Soares mous	1.12e-170
	24	104	5.5	509	9	AA177425	mt23d11.r1 Soares mous	1.12e-170
	25	101	5.3	331	11	AA332541	EST36505 Embryo, 8 wee	1.03e-163
	26	101	5.3	338	11	AA337465	oz88g03.s1 Soares_tota	2.13e-161
	27	100	5.3	343	21	AA992542	EST42223 Endometrial t	1.03e-163
	28	98	5.2	452	9	AA196933	zq59h08.r1 Stratagene	8.94e-157
	29	96	5.1	321	19	FI1594	HSC22A101 normalizd i	3.64e-152
	30	94	5.0	358	32	N36726	YX90c03.r1 Soares mela	1.44e-147
	31	95	5.0	360	35	CI3265	CI3265 Yuji Kohara unp	7.27e-150
	32	94	5.0	455	13	AA427173	ve83c01.r1 Soares mous	1.44e-147
c	33	95	5.0	506	26	AA901227	UI-R-AI-gp-g-06-0-UI.s	7.27e-150
	34	95	5.0	642	13	C23820	C23820 Dictyostellium d	7.27e-150
	35	93	4.9	365	11	AA325019	EST27941 Cerebellum II	2.83e-145
	36	93	4.9	704	9	AA192035	rs04h04.r1 Sommer Pris	2.83e-145
c	37	93	4.9	752	25	AI327328	mp70f10.r1 Soares 2NbM	2.83e-145
	38	90	4.8	302	12	AA356125	EST64527 Jurkat T-cell	2.05e-138
	39	90	4.8	461	9	AA196934	zq59h09.r1 Stratagene	2.05e-138
	40	90	4.8	568	24	C96654	C96654 Rice callus Ory	2.05e-138
	41	88	4.7	191	34	W49861	zEST00324 Maize Leaf,	7.34e-134
	42	88	4.7	331	16	AA682985	ae89c01.s1 Stratagene	7.34e-134
	43	88	4.7	337	11	AA321974	EST24544 Cerebellum II	7.34e-134
	44	86	4.5	490	8	M88895	CEH10F11 Chris Martin	2.54e-129
	45	85	4.5	510	21	C88687	C88687 Mouse early bla	4.65e-127

ALIGNMENTS

RESULT 1 AA896953 471 bp mRNA EST 06-APR-1998
 LOCUS L30-386T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
 DEFINITION NAC1 treatment Mesembryanthemum crystallinum cDNA clone L30-386 5'
 similar to Citrate synthase, mitochondrial precursor, mRNA
 sequence.
 ACCESSION AA896953
 NID 93033346
 VERSION AA896953.1 GI:30333346
 KEYWORDS EST.
 SOURCE common ice plant.
 ORGANISM Mesembryanthemum crystallinum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Caryophyllales; Caryophyllales; Alzooceae; Mesembryanthemum.
 REFERENCE 1 (bases 1 to 471)
 AUTHORS Cushman, J.C.
 TITLE An expressed sequence tag database for the common ice plant,
 Mesembryanthemum crystallinum
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150267.

Contact: Cushman JC
 Department of Biochemistry and Molecular Biology
 Oklahoma State University
 350 Noble Research Center, Stillwater, OK 74078-3035, USA
 Tel: 405-744-6207
 Fax: 405-744-7799
 Email: jcushman@biochem.okstate.edu
 PCR Primers
 FORWARD: T7
 BACKWARD: T3
 Plate: L30-4, row: H column: 6
 Seq primer: T3

High quality sequence stop: 320.

FEATURES

Location/Qualifiers

source

1..471

/organism="Mesembryanthemum crystallinum"

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1

EcoRI; Site_2: XhoI"

/db_xref="taxon:3544"

/clone="L30-386"

/clone_lib="Ice plant Lambda Uni-Zap XR expression

library, 30 hours NaCl treatment"

/tissue_type="Leaf, 30 h 0.4M NaCl"

/dev_stage="Six week old"

127 a 95 c 122 g 127 t

BASE COUNT

ORIGIN

Query Match 12.7%; Score 241; DB 20; Length 471;

Best Local Similarity 79.6%; Pred. No. 0.00e+00;

Matches 351; Conservative 0; Mismatches 86; Indels 4; Gaps 4;

Db 3 CATTGGTGGGAGGACCTTTACAGATCTTACCTTTTCATTGCGAGTGCACATCAATGGG 62

QY 922 CACTTGGTTGCTAGTCTTGTCTATCTTACCTCTCTTTGCTGCTTTGAAAGGT 981

Db 63 TTGGCTGGCCCACTTCATGGATTAGCCAAATCAGGAAGTGTGCTGTGATCAAGTCAAGT 122

QY 982 TTAGCCGGACCACTTCATGGTTTAGCCAAATCAGGAAGTGTGCTGTGATCAAGTCAAGT 1041

Db 123 GTGGATGAATGTGGAGAAACATCAGCAGGAGAGCTTAAAGATTATGCTGGAAGACA 182

QY 1042 GTAGAAGATGTGGGAGAACATTTCCAAAGAGAGCTTGAAGACTATGTTTGGAAACA 1101

Db 183 TTAAGAGTGGCAAGTGTGCTGATATGGCCATGGAGTCTCGCAGACGGATCCA 242

QY 1102 TTGAACAGTGGCAAGTGTGCTGCTGTTTGGACATGAGTCTCGCAGACGGATCCA 1161

Db 243 AGATATCTTGGCAGAGGACCTTTGCCCTTAAGCACTTACCCTATGATCCACTATTCAA 302

QY 1162 AGATATACATGCGAGAGAGTTCGCTATGAAGCACTTTCGCTGAAGATCCACTGTTCAA 1221

Db 303 CTGGTTTGAAGCTGTATGATGGTGGCTTCCCAATTTCTCAGAGCTTGGGAAGTAAAGA 362

QY 1222 CTGGTTTCAAACTCTACGAAGTCTTCTGCTGCTTCTACAGACTTGGCAGAGTAA 1280

Db 363 ACCATGGCCGAATGTGATGCGCCAGTGGGTGATGCTGAACACTATGATGTTGACAA 422

QY 1281 ACCTTGGCCAAATGTGATGTCGCCACAGTGG-TGTGTTGTTGAACATATTATGTTAACT- 1330

Db 423 GAAAGCAAGATACACTGT 443

QY 1339 GAA-GCAAGATATTACGGT 1358

RESULT 2 T44185 502 bp mRNA EST 07-JAN-1998

LOCUS 7448 Lambda-PRL2 Arabidopsis thaliana cDNA clone :23C917, mRNA

DEFINITION sequence.

T44185

NID 92758988

VERSION T44185.1 GI:2758988

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

Rosidae; Capparales; Brassicaceae; Arabidopsis.

1 (bases 1 to 502)

Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,

McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M.,

Retzel, E., and Somerville, C.

Genes galore: a summary of methods for accessing results from

large-scale partial sequencing of anonymous Arabidopsis cDNA clones

Plant Physiol. 106, 1241-1255 (1994)

95148729

MEDLINE

COMMENT

On Jan 7, 1998 this sequence version replaced gi:948518.

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tcn@ibm.cl.msu.edu

Seq primer: T7.

Location/Qualifiers

1..502

source

/organism="Arabidopsis thaliana"

/strain="var columbia"

/note="Vector: lambda Zip-Lox. Site_1: Sal, Site_2: Not.

Lambda PRL2 is a cDNA library derived from equal

quantities of 4 pools of mRNA. The mRNA sources were 1) 7

day germinated etiolated seedlings; 2) tissue culture

grown roots; 3) staged plants half with 24 hour light

cycle, half on 16 hr light, 8 hour dark-rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and

siliques. The vector is BRL's lambda Zip-Lox. The cDNA

inserts were directionally cloned with Sal-Not arms using

oligo dt primed cDNA.

/db_xref="taxon:3702"

/clone="123C97"

/clone_lib="Lambda-PRL2"

BASE COUNT 125 a 90 c 118 g 149 t 20 others

ORIGIN

Query Match 10.6%; Score 200; DB 17; Length 502;

Best Local Similarity 79.1%; Pred. No. 0.00e+00;

Matches 291; Conservative 0; Mismatches 74; Indels 3; Gaps 2;

Db 2 AATCTTTGGATTATGGTGCAAAATTTTCCACATGTTGGGATTGAT-GAT--GAAAAGG 58

QY 788 AATCCCTGGATTATGGTGCAAAATTTTGTCTACATGCTTGGTTTCAGTAGCTCTCAATGC 847

Db 59 TGAAGAGCTCATGAGGCTTTACATCACCATCCACATGATCATGAAGTGGAAATGTTA 118

QY 848 ATGAATCTTCTATGAGGCTCTATGTAAATACACATGATCATGAAGTGGTAAATGTC 907

Db 119 GTGCTCACACTGGTACCTGGTGTGTAGTGCACATTCAGACCCATATCTGTCAATTTGCAG 178

QY 908 GTGCTCACACCGGTACCTGGTGTGTAGTGTCTGTCTGATCTCTACCTCTCTCTGCTG 967

Db 179 CTGCATTAAATGGTTTAGCTGGGCCACTCCATGGTTTGGCTATATCAGGAAGTTTGTCTT 238

QY 968 CTGCTTGAATGGTTTAGCGCGACCACTTTCATGGTTTGGCCAAATCAGGAAGTTTGTCTAT 1027

Db 239 GGATCAAAATCAGTGTAGAGGAATGTGGAGAGATATATCAAAAGAACAGCTTCAGGAAGAT 298

QY 1028 GGATAAATCTGTGTAGAGAAATGTGGGAGAACATTTCCAAAGACAGCTTCAGAGACT 1087

Db 299 ATGTTTGGGAAACATTAACAGTGGCAAGGTATTCCGGGGTATGTCACGGTGTCTTCNC 358

QY 1088 ATGTTTGGGAAACATTAACAGTGGCAAGGTATTCCCTGGTTTGGACATGGAGTTCTGC 1147

Db 359 GCAATACT 366

QY 1148 GAAAGACT 1155

RESULT 3

LOCUS R90544 478 bp mRNA EST 30-DEC-1997

DEFINITION 16899 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189D1577, mRNA

sequence.

R90544

NID g958084

VERSION R90544.1 GI:958084

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparidales; Brassicaceae; Arabidopsis.
1 (bases 1 to 478)
Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
Retzel, E. and Somerville, C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
On Apr 14, 1993 this sequence version replaced gi:693655.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.

FEATURES
source
1..478
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="var columbia"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dI primed cDNA."
/db_xref="taxon:3702"
/clone="189D157"
/clone_lib="Lambda-PRL2"
BASE COUNT 119 a 90 c 113 g 137 t 19 others
ORIGIN

Query Match 10.0%; Score 189; DB 31; Length 478;
Best Local Similarity 79.6%; Pred. No. 0.00e+00;
Matches 256; Conservative 0; Mismatches 66; Indels 2; Gaps 2;
Db 4 TCAGACCATATCTGTCATTTGAGCTGCATTAAATGGTTAGCTGGCCACTCCATGCT 63
QY 943 TCTGATCTTACCTCTCTTTGTTGCTGCTTTTGAATGGTTAGCTGGACCACTTCATGCT 1002
Db 64 TTGGCTAATCAGGAAGTTTTCCTTTGGATCAATCAGTCGTAGAGGAATGTGGAGAAGAT 123
QY 1003 TTAGCCAATCAGGAAGTTTTCCTATGATAAATCTGTTAGAACAAATGTGGGAGAAC 1062
Db 124 ATATCAAAAGACAGTTGAAGATATGTTTGGAAACATTAAACAGTGGCAAGTTAT 183
QY 1063 ATTTCAAAAGACAGTTGAAGACTATGTTTGGAAACATTGAACAGTGGCAAGTTGTC 1122
Db 184 CCGGATATGCTCAGCTGTTCTCGCAATACTGATCCAGATATGATGCCAAGAGAA 243
QY 1123 CTTGGTTTGGACATGGAGTTCTGGAAACACGTACCAAGATATACATCCAGAGAG 1182
Db 244 TTGGCTTGAAGCATCTACCTACGCGCTCTTTTCCAGCTGGTGTCAAAGCTTTATGAA 303
QY 1183 TTCCTATGAAGCATTTGCTCGTGAAGTCCACTGTTTCAACTGGTTTCAAACCTCAGAA 1242
Db 304 GTTGTGCTCCCTGCTTCTCACTGAGCTTTGGAAA 337
QY 1243 GTTTT-CCTCC-TGTTTCTACAGAACTTGGCAA 1274

RESULT 4

LOCUS AU002699 620 bp mRNA EST 15-JAN-1999
DEFINITION AU002699 Bombyx mori p50(Daizo) Bombyx mori cDNA clone n0357, mRNA
sequence.
ACCESSION AU002699
NID AU002699
VERSION AU002699.1 GI:4159092
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Lepidoptera; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 620)
AUTHORS Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsa@nirs.go.jp
PROJECT - 'CREST Project by JST'.

FEATURES
source
1..620
Location/Qualifiers
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="n0357"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT 162 a 141 c 146 g 171 t
ORIGIN

Query Match 9.5%; Score 179; DB 26; Length 620;
Best Local Similarity 66.4%; Pred. No. 0.00e+00;
Matches 392; Conservative 0; Mismatches 195; Indels 3; Gaps 3;
Db 1 CACCACGAGGTGGAAAGCTTTCGCGCACACTACGCAATTTAGTGGTTCTGCTCTTAGT 60
QY 886 GATCATGAAGTGGTATGTCAGTCTCACACCGTCTCACTGTTGTAGTCTTGTCT 945
Db 61 GATCCTTACTTATCTTCGCGCTGGACTCAATGGTCTTGTGGACCACTTCAGGCTG 120
QY 946 GATCCTTACTCTCTTGTGCTGCTTGAATGGTTAGCGGACCACTTCATGGTTA 1005
Db 121 GCTAATCAGAGGTTCGTGATGCTCGAGAACTGCGCAAACTGGCAAACTTGTGATAATTC 180
QY 1006 GCCATCAGGAAGTTTGTCTATGATAAATCTGTTAGAGAATGTGGGAGAACATT 1065
Db 181 ACAGAAGAGCAACTCAAGAATTCATCTGGAACACACTTAAATCTGTGAGTTGTACCT 240
QY 1066 TCCAAAGAGCAGTTGAAAGACTATGTTTGGAAACATTTGAACAGTGGCAAGTTGCTCT 1125
Db 241 GCTTACGCTCATGAGTACTTAGAAACACTGATCCAGATATACTTCCGCGGTGAGTTT 300
QY 1126 GTTTTGGACATGAGTTCTCGGAAAGACTGTACCAAGATATACATCCAGAGAGATT 1185
Db 301 GCTCTTAAGCACTTACCCCAATGACCAATTTCAAGTGGTGGCTGTGTTTACAGGTT 360
QY 1186 GCTATGAAGCATTTGGCTGAAGATCCACTGTTTCAACTGGTTTCAAACTCTACGAAGTT 1245
Db 361 GTTCCGCGATCTCCACCGAAGTGGCAAGTCAAGAACCCATGGCCTATGTAGACTCC 420
QY 1246 -TTCTCTCTGTTTCTACAGAACTTGGCAAGTTAA-AACCT-TGGCCAAATGTTGATGCC 1302
Db 421 CATTCGGGAGTCTTTTTCAGTATTATGGTCTGAAGAGATGAACCTATATACAGTAATG 480
QY 1303 CACAGTGGTGTGTTGTTGAATATATGTTTAACTGAAGCAAGATATATACGGTCTC 1362
Db 481 TTTGGAGTGTCCGAGCTTGGGTGTTCTCGCTCAGTTGATTGGTCCCGTCCGCGA 540
QY 1363 TTTGGCGTATCAAGAGCTCTTGGCATTTGCTCTCAGCTAATTTGGGACCGAGCTTTGGA 1422

```

Db 541 CTCCCATTCAGAGACCAAAAGTCGCTCAGCAGGAGATGCTTATGAACA 590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1423 TTCCCGCTAGAGAGCCAAAGAGCTGTCAACATGGAGTGGCTTGAGAACCA 1472

RESULT 5
LOCUS C96653 417 bp mRNA EST 19-OCT-1998
DEFINITION C96653 Rice callus Oryza sativa cDNA clone C10531_5A, mRNA
sequence.
ACCESSION C96653
NID g3758954
VERSION C96653.1 GI:3758954
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 417)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'
Location/Qualifiers
1. 417
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/notes="Vector: pBluescript II SK+; Site.1: SalI; Site.2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
/db_xref="taxon:4530"
/map="11"
/clone="C10531_5A"
/clone_lib="Rice callus"
BASE COUNT 115 a 76 c 108 g 118 t
ORIGIN

Query Match 9.1%; Score 172; DB 24; Length 417;
Best Local Similarity 73.1%; Pred. No. 0.00e+00;
Matches 304; Conservative 0; Mismatches 108; Indels 4; Gaps 3;

Db 2 TATGTTTACGGAGGATCTTCAAGATGGAAACATAGCAGCTGATATGCACCTGGAC 51
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 739 TATGTTTATCGCAGGATGTACAGAAATGGTGACATATACCTTAAGGATGAATCCCTGGAT 798

Db 62 TATCGAGCAAACTTTTACACATGTTTGGGTTTGATGATCCCAAAATGCTCGAG--T-TG 118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 799 TATGGTGCAAAATTTTCTGCATCATGCTTGGTTTCAGTAGCTCTGAAATGCATGAATCTTT 858

Db 119 ATCGCACTATATGTGACAAATCCACACTGATCATGAAGTGGAAATGTGAGTGTCTACT 178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 859 ATAGGCTCTATGTAACAATACACAGTATCATGAAGTGGTAAATGTGAGTGTCTCAGCACC 918

Db 179 GGACATCTGTTGGAAGTCTCTGTGCAGATCTTATCTTTCTTTTTCAGCTGCACTGAAT 238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 919 GGTCACTTGGTGTCTAGTCTTTGTGATCTTACCTCTCCTTGTGCTGCTTTGAAT 978

Db 239 GGTTTAGCTGGACCATTCGACGCTGGTGAATATGAGGAAGTGTCTGTCTGATCAAAATCT 298
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 979 GGTTTAGCGGGACCATTCATGGTTTAGTCAATCAGGAAGTTTGTCTATGATGAATAATCT 1038

```

```

Db 299 GTAATAGGTGACAGAGTAGTGTATGTTTACAACTGATCAACTCAAGAGTATGTTGGAAG 358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1039 GTGTGAGAAGAATGTTGGGAGAACAAITTTCCAAAGAGCAGTTGAAAGACTATGTTTGAAA 1098

Db 359 ACACATAAAGTGGAAAGTGTGTTCTCTCGCTC-GGTGATGGAGTTCTACGTAAAGAC 413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1099 ACATTCGAACAGTGGCAAGGTGTCCCTGGTTTGGACATGAGGTTCTGCGAAGAC 1154

RESULT 6
LOCUS A1176862 676 bp mRNA EST 20-JAN-1999
DEFINITION EST220465 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROV8X80 3' end, mRNA sequence.
ACCESSION A1176862
NID g3727500
VERSION A1176862.1 GI:3727500
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 676)
AUTHORS Eutheria; Rodentia, Sciurognathi, Muridae, Murinae, Rattus.
TITLE Kerlavage,A.R. and Adams,M.D.
JOURNAL Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152042.

Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. 676
/organism="Rattus sp."
/notes="Organ: Ovary; Vector: pT73Pac; Site.1: EcoRI;
Site.2: NotI"
/db_xref="ATCC (inhost):2031570"
/db_xref="taxon:10118"
/clone="ROV8X80"
/clone_lib="Normalized rat ovary, Bento Soares"
BASE COUNT 156 a 188 c 150 g 181 t 1 others
ORIGIN

Query Match 9.1%; Score 173; DB 23; Length 676;
Best Local Similarity 67.8%; Pred. No. 0.00e+00;
Matches 375; Conservative 0; Mismatches 173; Indels 5; Gaps 5;

Db 123 CTTCGGCTTTCTAGAGGAAGCCTAGGCTCTGCTCCAGATAAGCTGGGCCAACACACC 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1443 CTTTGGCTCTCTAGCGCAATCCAAAGAGCTCGGTCCCAATTAAGTGAAGCAAA180C 1384

Db 183 CAGTGGCTTCGACACTCGGACAGCACTGTCTAGTAAATCATCTCCCTATGCCATAGTA 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1383 AGAGCTCTTGATAGCCCAAGAGGACCGGTATAATATCTTCTTCAGTTAAACCATATA 1324

Db 243 CTGGAGAGACGCCACCTGTGAGCATCTACGTTGGGCGAGGGTCTTACGCTTCTCTTG 302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1323 GTTCAACACACACACCTGTGGGATCAACATTTGGTCAAGG-TT-TTAAATTGGCAAG 1266

Db 303 CTTCAAGAGGATTTGGCCAAATCTTGTACAGCTGAGCAGCAGCTTAACATGGCATC 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1265 TTCTGTAGACAGAGAGAAACT-TGTAGAGTTTGAACACAGTTGAACACAGTGGATC 1207

Db 363 CTTGGGCAAGATTTTTCAGAGCAAACTCTCGGTGACAGGAATATCGTGATCACTTCTCT 422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1206 TTCAGGCAAAATGCTTCATAGGCAACTCTCTCTGGCATGTATATCTTGTGTACAGTCTTTCG 1147

```



```

Db 423 TAGAGTCGATGACCATATCTCTGGGACAAACCCGTCOCGAGTTGAGTGTCTTCCAGATGTA 482
Cp 1146 CAGAACTCCATGTCCTCAAAACAGGAGCAACCTTGCCACTGTTCAATGTTTCCAAACATA 1087
Db 483 GTCTCGTAACCTTCTCATCTGACAGCTCTTTGGCGACTTCTCTGTAGCTGGTC-AGCC 541
Cp 1086 GTCTTTCAACTGCTCTTTGGAAATGT-TCTCCCAACATCTCTCTACAACAGATTTATCC 1028
Db 542 AGACAAGCAGCTCTCTGTTTGTAGTCCATGTCAGAGGCGCTGCGACGCCCATTCATAGCTG 601
Cp 1027 ATAGCAAACTTCTGATGCTTAACCATGAAGTGGTCGGCTAAACCATTCACAAGCAG 968
Db 602 CTCANAGGACAGTAAAGGCTCTGAAGGCGCTGCCACCAAGTGGCTGTGTGGGAC 661
Cp 967 CAGCAAGGAGAGGTAAAGATCAGACAAAGCACTAGCAACCAAGTGACCGGTGTGAGCAC 908
Db 662 TTACATTGCCACC 674
Cp 907 TGACATTACCACC 895

7
RESULT AA606966 657 bp mRNA EST 30-SEP-1997
LOCUS vm94h09.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
DEFINITION clone IMAGE:1005953 5' similar to SW:CI5Y_PIG P00889 CITRATE
SYNTHASE, MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION AA606966
NID 92455859
VERSION AA606966.1 GI:2455859
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 657)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, I., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1328460.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:570169
High quality sequence stop: 473.
FEATURES
source Location/Qualifiers
1..657
/organism="Mus musculus"
/strain="B6D2 FL/J"
/notes="Organ: embryo; Vector: pSPORT; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGTCCAGCTCGACCGCTTTTGT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (large-
inserts) and B3."
/db_xref="taxon:10090"
/clone="IMAGE:1005953"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
BASE COUNT 161 a 167 c 181 g 148 t

```

```

ORIGIN
Query Match 8.1%; Score 153; DB 15; Length 657;
Best Local Similarity 66.7%; Pred. No. 3 25e-287;
Matches 396; Conservative 0; Mismatches 189; Indels 9; Gaps 9;
Db 22 CTCATGGCTTTGTACCTCACCATCCATAGTACCATGAGGCTGCTAATGTAAAGTCCAC 81
Qy 856 CTTATGAGGCTCTATGTAAACAATACAGATGATCATGAAGTGGTAAATGTCAAGTGCAC 915
Db 82 ACAAGCCATTTGGTGGCAGCGCAAT-TCAGACCCCTTACCTGTCTCTTGCAGACGCCATG 140
Qy 916 ACCGCTCACTTGGTGTAGTGTCTTGTCTGATCCTTACCTCTCTTGTGTGCTGTG 975
Db 141 AATGGCTGGCGGGGCT-CTACATGGACTAGCAAAATCAGGAGTGCTTGTCTGCTGACA 199
Qy 976 AATGGTTAGCCGACCCACCTTCATGGTTTACCCCAATCAGGAAGTTTGTATGATAAAA 1035
Db 200 -CAGCTACAGAAGGAAGTTGGCAAAGAGCTGTGATGAGAGTTAGAGACTACATCTG 258
Qy 1036 TCTGTGTAGAAAGATGTGGGAGA-ACATTTCCAAAGAGCAGTTGAAAGACTATGTTG 1094
Db 259 GAACACACTCAATTCAGACGCGTGGTCCAGGATACGGTCATCAGTACTGAGGAAGAC 318
Qy 1095 GAAACATTTGAACAGTGGCAAGGTTGCCCTGGTGGTGGACATGGAGTTCTGCCAAGAC 1154
Db 319 TGACCTCGCTATTCCTGTACGCGAGAGTGTGCTCTGAAACATCTGCTAAGGATTCCT 378
Qy 1155 TGTACCAAGATATACATGCCAGAGAGAGTTCGCTATGAAGCATTTGCTGAAG-ATCCAC 1213
Db 379 ATGTTCAAGCTGGTGGCTCAGCTGTACAAGATTTGCCCAATATCCTCTTAGACAAGG 438
Qy 1214 TGTTCAT-CTGGTTTCAAACTCTACGAGTTTTCCTCC-TGTTCTTACAGAACTTGGC 1271
Db 439 AAGCTAAGAACCTTGGCCCAAGTAGACGCTCACAGTGGGTGCTGCTCCACTACTAT 498
Qy 1272 AAAGTTAA-ACCT-TTGCCCAATTTGTATGCCACAGTGGTGTGTTGTTGAATATAT 1329
Db 499 GGCATGACGGAGATGAAGTACTACACAGTCTCTGTTGGAGTGTCTCGGGCACTGGGTG 558
Qy 1330 GGTTTAACTGAAGCAAGATATTATACGGTCTCTTTGGCGTATCAAGAGCTCTTGGCAT 1389
Db 559 CTAGCCAGCTCATCTGTGAGCAGAGCCCTAGGCTTCCCTTGGAAAGCCCAAG 612
Qy 1390 TGCTCTCAGCTAATTTGGACCGAGCTCTTGATTTGGCTTAGAGAGCCCAAG 1443

RESULT 8 AA231705 307 bp mRNA EST 27-FEB-1997
LOCUS CDO534.R cDNA from oat Avena sativa cDNA clone CDO534, mRNA
DEFINITION sequence.
ACCESSION AA231705
NID g1854091
VERSION AA231705.1 GI:1854091
KEYWORDS EST.
SOURCE oat.
ORGANISM Avena sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Avena.
REFERENCE 1 (bases 1 to 307)
AUTHORS Vandeynze, A.E., Sorrells, M.E., Park, W.D., Ayres, N.M., Fu, H.,
Cardenhour, S.W. and McCouch, S.R.
TITLE Anchor Probes for Comparative Mapping of Grass Genera
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:634857.
Contact: McCouch SR
Dept Plant Breeding
Cornell University
Ithaca, NY 14853-1901, USA
Tel: 607 255 0420
Fax: 607 255 6683

```

FEATURES source
 Email: srm4@cornell.edu
 CDNA from oat (Avena sativa); reverse sequence of RFLP probe
 CDO534. Sequence determined by Nicola M. Ayres.
 Location/Qualifiers
 1..307
 /organism="Avena sativa"
 /cultivar="Brooks"
 /note="Vector: Uni-ZAP XR/pBluescript; Site 1: EcoRI;
 Site 2: XhoI; A Uni-ZAP XR cDNA library was constructed
 from etiolated leaf mRNA from the oat cultivar 'Brooks'
 and converted to pBluescript (amp resistant) as described
 in Heun et al. (1991) Genome 34:437-447. For insert
 amplification, use M13 forward and reverse primers.
 Clones from this library are designated with the prefix
 'CDO'. *Note: Clone CDO1081 was recloned into the TA
 cloning vector and carries kanamycin resistance."
 /db_xref="taxon:4498"
 /clone="CDO534"
 /clone_lib="cDNA from oat"
 /tissue_type="etiolated leaf"
 84 a 64 c 79 g 80 t

Query Match 7.9%; Score 150; DB 10; Length 307;
 Best Local Similarity 75.2%; Pred. No. 5,79e-280;
 Matches 231; Conservative 0; Mismatches 75; Indels 1; Gaps 1;
 Db 1 GCGCAGCTGAACGGTTTAGCTGGCCACTGCATGGCTGGCTTAATCAGGAAGTGTGCTA 60
 QY 967 GTCTGTTGAATGGTTTACGGCGACCACTCATGTTTAGCCCAATCAGGAAGTGTGCTA 1026
 Db 61 TGGATCAAAATCTCTGATGGAAGAAACCGGAGTAAATACAACTGATCAGCTTAAAGAA 120
 QY 1027 TGGATAAAATCTCTTCTGAAGAAGATGCGGGAGAACATTTCCAAAGAGCAGTTGAAGAC 1086
 Db 121 TAGTATGGAAGACACTGAAGAGTGAAGAGTTGTTCTCGGTATGTCATGGAGTTCTA 180
 QY 1087 TAGTATGGAAGACACTGAAGAGTGAAGAGTTGTTCTCGGTATGTCATGGAGTTCTG 1146
 Db 181 CGTAATACAGATCCAGCATCTCGTCCCAAGGAGTTTGCACCTTAAGTATTACCCGAG 240
 QY 1147 CGAAGACTTACCAAGATATACATGCCAGAGAGATTCGCTATGAAGCAATTCGCTGAA 1206
 Db 241 GACCCACTTTTCAACTGGTCTCCAAAGTTGTAGGAAGTTGGCCCTCTCACTTACTGAG 300
 QY 1207 GATCCACTGTTTCAACTGGTCTCCAAAGTTGTAGGAAGTTGGCCCTCTCACTTACTGAG 1265
 Db 301 TTAGGCA 307
 QY 1266 CTGGCA 1272

RESULT 9
 LOCUS R90561 527 bp mRNA EST 30-DEC-1997
 DEFINITION 16916 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189H317, mRNA
 sequence.

ACCESSION R90561
 NID 9958101
 VERSION R90561.1 GI:958101
 KEYWORDS EST.

SOURCE thale cress,
 Arabidopsis thaliana

ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Rosidae; Capparales; Brassicaceae; Arabidopsis.

REFERENCE
 1 (bases 1 to 527)
 Newman, T., deRuijin, F.J., Green, P., Keegstra, K., Kende, H.,
 McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
 Retzel, E., and Somerville, C.

TITLE
 Genes galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 JOURNAL Plant Physiol. 106, 1241-1255 (1994)
 MEDLINE 95148729

COMMENT

On Apr 14, 1993 this sequence version replaced gi:593674.

Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313t@msu.edu
 Seq primer: T7 dye primer.

FEATURES

Location/Qualifiers
 1..527

source
 /organism="Arabidopsis thaliana"
 /strain="var columbia"
 /note="Vector: lambda zip-Lox; Site 1: Sal; Site 2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques). The vector is BRL's lambda zip-Lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dt primed cDNA."
 /db_xref="taxon:3702"
 /map="873F09; 1q21.3-1q23.2; 9"
 /clone="189H37"
 /clone_lib="Lambda-PRL2"
 130 a 106 c 122 g 148 t 21 others

BASE COUNT
 ORIGIN

Query Match 7.9%; Score 150; DB 31; Length 527;

Best Local Similarity 73.3%; Pred. No. 5,79e-280;

Matches 321; Conservative 0; Mismatches 105; Indels 12; Gaps 12;

Db 3 ATTCTGAGTGCAGAAAGTATTACCTACTCCAGCTGAGAGTGAACCAATTACCGGAG 62

QY 385 ATACCTGAATGCCAAAGGTATTACCTGCAGAAAGCGTGGGGTGGAGCCCTGCCTGAA 444

Db 63 GGTCTTTGTGGCTTCTTACTGGAAGGTACTAGCAAGAGCAAGTGAAGCACT- 121

QY 445 GGTCTTCTCTGGCTTCTTTTACAGGAAGGTGGCCATCAAAGAGCAAGTGAATTCATT 504

Db 122 GTCGAAGACTTGGGAAACCGTGTGC-TGTGCCAGATTATG-TGTCAATGCCATCGAT 179

QY 505 GTCTCAGGAATTGCAGAGTCGGGCATCATATCCCTGATCATCATATACAACTATTGAT 564

Db 180 GCTCTGCCCTTCACAGCTCATCCAAATGACTCAATTTGCTAGCGGTGTTATGGCCCTCCAG 239

QY 565 GCCTTACCAGTCACAGCTCATCCAAATGACCCAGTTTGTACTGGAGTCATGGCTTTCAG 624

Db 240 GTGCAAGTGAAGTTCACAAAGGCATATGGAATGGAATTCATAAGTCAAAAGTCTCGGAG 299

QY 625 GTTCAAGTGAATTTCAAAGGCATACGAGAAAGGATTCAAAATCAAGTATTGGAA 684

Db 300 CCAACATATGAGGATTGGCTCAACCTGATTTGCTCGTGTTCCTGTTGAGNCTGCAATA 359

QY 685 CCAACATATGAGGATTGCCAATCTGATT-GCTCAAGTT-CCACTTGTG-CTGCTTA- 740

Db 360 TGTATTATCGAAGGATGTATTAANAATTTGGTTCCTATCCCTCAGGATAAAACCTTTG 419

QY 741 TGTATTATC-GCAGG-ATGTACAAGAAAT-GGTGACACTATACC-TAAGGATGAATCCCT-G 795

Db 420 GNTTNTGGNNNAATTT 437

QY 796 GATTATGGTGAATTTT 813

RESULT 10

LOCUS AA313713

DEFINITION EST185581 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end

similar to similar to citrate synthase, mRNA sequence.

EST 19-APR-1997


```

Bonaldo."
/db_xref="taxon:10090"
/map="3"
/clone="IMAGE:679873"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"

BASE COUNT 129 a 125 c 150 g 118 t
ORIGIN

Query Match 7.2%; Score 136; DB 10; Length 522;
Best Local Similarity 66.9%; Pred. No. 2,69e-246;
Matches 349; Conservative 0; Mismatches 165; Indels 8; Gaps 8;

Db 1 ATGAGGCTGTAATGTAAGTGGCCACACAGCCATTTGGTGG-AGCCTGT-TCAGACC 58
|||||
Qy 890 ATGAAGGTGTAATGTAAGTGGTCTACACCGGTCACATTTGCTAGTCTTTGCTGATC 949
|||||

Db 59 CTTACCTGTCTTTGACGAGCCATGAATGGGCTGCGGGCT-CTACATGGACTAGCAA 117
|||||
Qy 950 CTTACCTGTCTTTGCTGCTTTGAATGGTTAGCCGGACCACTTCATGTTAGCCA 1009
|||||

Db 118 ATCAGGAGTGTCTGTGCTGACA-CAGCTACAGAGGAAGTTGGCAAGACGTGTCA 176
|||||
Qy 1010 ATCAGGAAGTTTGCATGATAAAATCTGTTAGAGAATGTGGGAGA-ACATTTCC 1066
|||||

Db 177 GATGAGAAGTTACGAGACTACATCTGSAACACACTCAATTCAGAGCGGTGTCACGGA 236
|||||
Qy 1069 AAGAGCAGTTGAAGACTATGTTGGAAACATTTGAACAGTGGCAAGGTTGTCCTGGT 1128
|||||

Db 237 TAGCGTACGACTACTGAGGAAGACTGACCCCTCGCTATTCCTGTACGAGAGTTTCT 296
|||||
Qy 1129 TTTGGACATGGAGTTCTGCGAAGACTGTACCAAGATATACATGCCAGAGAGTTTCT 1186
|||||

Db 297 CTGAACATCTGCTAAGATCCATGTTCAAGCTGGTGGCTACGCTGTACAGATTGTG 356
|||||
Qy 1189 ATGAAGCATTTGCTGAAGATCCACTGTTCAACTGGTTTCAAACTCTACGAAGTTTC 1248
|||||

Db 357 CCCAATATCTCTTAGAGCAAGGAAGGCTAAGAACCTTTGCCCAAGCTAGACGCTCAC 416
|||||
Qy 1249 CTCC-TGTTCTTACAGACTTGCAAAAGTTAA-AACC-TTGCCAAATGTTGATGCCAC 1305
|||||

Db 417 AGTGGGTGCTGCTCAGTACTATGCGATGAGGAGATGAATATACACAGTCTCTGTT 476
|||||
Qy 1306 AGTGGTGTGTTGTAATATTATGTTTAACTGAAGCAAGATATTATACGCTCTCTTT 1365
|||||

Db 477 GGAGTGTCTCGGCACTGGGTGCTAGCCGACGCTCATCTGG 518
|||||
Qy 1366 GCGGTATCAAGACTCTTGCACTTTGCTCTCAGCTAAITGG 1407
|||||

RESULT 12
LOCUS 217455 213 bp mRNA EST 10-NOV-1992
DEFINITION ATTS0029 AC16H Arabidopsis thaliana cDNA clone TAT2B1 5' similar to CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR Swiss-Prot entry P00889, mRNA sequence
ACCESSION 217455
NID 217455
VERSION 217455.1 GI:16597
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 213)
AUTHORS CNRS
TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program
JOURNAL Unpublished (1996)
COMMENT Contact: Bardet C., Axelos M., Tremousaygue D., Lebas M., Lagrave T., Lescure B.
UMR05 CNRS/INRA

laboratoire de Biologie Molculaire
BP 27,31326 Castanet-Tolosan cedex,France
Email: lescure@toulouse.inra.fr.
Location/Qualifiers
1..213
/organism="Arabidopsis thaliana"
/strain="ecotype Columbia"
/note="Vector: Lambda ZAPII; tissue=cell suspension culture of ecotype columbia; clone_library=AC16H; Cloning vector: Lambda ZAPII; Physiological condition: cycling cells."
/db_xref="taxon:3702"
/clone="TAT2B1"
/clone_lib="AC16H"

BASE COUNT 59 a 40 c 52 g 62 t
ORIGIN

Query Match 6.6%; Score 125; DB 8; Length 213;
Best Local Similarity 82.3%; Pred. No. 4,57e-220;
Matches 172; Conservative 0; Mismatches 35; Indels 2; Gaps 2;

Db 7 ACATCCACAGTATCATGAAGTGGAAATGTTAGTCTCACACTGGTCACCTGGTTGGT 66
|||||
Qy 874 ACAATACACAGTATCATGAAGTGGTAAATGTCATGCTCACCGGTCACCTGGTTGCT 933
|||||

Db 67 AGTGCACTTTCAGACCCATATCTGTCATTGTCAGCTGCAATTAATGTTTAGCTGGCCA 126
|||||
Qy 934 AGTGCTTGTCTGATCTTACCTCTCCTGCTGCTGCTTGAATGTTTAGCCGGACCA 993
|||||

Db 127 CTCATGTTTGGCTAATCAGGAAGTTTGGCTTTGGATCAAT-TCAGTGTAGAGGAATGT 185
|||||
Qy 994 CTTCTGTTTAGCAATCAGGAAGTTTGGCTATGATGATAAAATCTGTTGAGAAGATGT 1053
|||||

Db 186 GGAGAAGATATATC-AAAGACAGTTGAA 213
|||||
Qy 1054 GGGGAGAACATTTCCAAAGACGAGTTGAA 1082
|||||

RESULT 13
LOCUS AAL12941 466 bp mRNA EST 01-AUG-1997
DEFINITION zn58b06.r1 Stragatene muscle 937209 Homo sapiens CDNA clone IMAGE:562451 5' similar to SW:CSI5_FIG P00889 CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION AAL12941
NID q1664291
VERSION AAL12941.1 GI:1664291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Pohlning,T., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,P.
WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:593136.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 2292 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 423.
Location/Qualifiers
1..466

```

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.
Location/Qualifiers
1. .252
/organism="Arabidopsis thaliana"
/strain="var columbia"
/note="vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1)
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dI primed cDNA."
/db_xref="taxon:3702"
/map="21q"
/clone="158K127"
/clone.lib="Lambda-PRL2"
57 a 54 c 64 g 67 t 10 others

BASE COUNT
ORIGIN

Query Match 6.1%; Score 115; DB 17; Length 252;
Best Local Similarity 79.5%; Pred. NO. 2.02e-196;
Matches 159; Conservative 0; Mismatches 39; Indels 2; Gaps 2;

Db 3 C T C T C T G T C T A C T G A C T T G G A A G G T G A A C C G T G G C C A A T N T T G A T G C T C A C 62
|||||
QY 1248 C T C T C T G T T T A C A G A A C T T G G C A A G T T A A - A A C T T - T G C C A A A T G T T A T G T G C C A C 1305
|||||

Db 63 A G T G G G G T T T G C T G A A C C A C T A T T G T T A A C C G A A G C A G G T A C T A C A C G T G C T C T T 122
|||||
QY 1306 A G T G G T G T T T G T T G A C T A T T A T T G T T A A C T A A G C A A G A C A T A T T A C G T C T C T T 1365
|||||

Db 123 G T G T T T C A A G A G T C T T G G N A T C N C T C A G C T T A T A T G G A C C A G C T C T T G A C T T 182
|||||
QY 1366 G C G C A T C A A G A G C T T T G G C A T T G C T C A G C T A A T T T G G A C C A G C T T T G G A T T G 1425
|||||

Db 183 G C A C T T G A N G G C C A A A G A G 202
|||||
QY 1426 C G C T A G A G A G C C C A A A G A G 1445
|||||

RESULT 15
LOCUS A1054878 585 bp mRNA EST 16-JUL-1998
DEFINITION coau0002G14 Cotton Boll Abscission Zone cDNA Library Gossypium
hirsutum cDNA clone coau0002G14 5', mRNA sequence.
ACCESSION A1054878
NID G3325992
VERSION A1054878.1 GI:3325992
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 585)
Wan, C.-H., Yu, Y., Sasinowski, M. and Wing, R.A.
Cotton EST Database: Sequence Analysis of 2000 cDNA Clones from an
Abscission Zone Library
Unpublished (1998)
JOURNAL On Jan 9, 1998 this sequence version replaced gi:930638.
COMMENT

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Wing RA
Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: SP030 (AACAGCTATGACCATGATTA)
High quality sequence stop: 239.

FEATURES

source
1...585
/organism="Gossypium hirsutum"
/note="Vector: Uni-ZAP XR; Site_1: EcoRI; Site_2: XhoI;
This is a Uni-ZAP XR custom cDNA library made by
Stratagene (U.S.A.: 1-800-424-5444): Stratagene cat.
#837201."
/db_xref="taxon:3635"
/clone="coau0002G14"
/clone_lib="Cotton Boll Abscission Zone cDNA Library"
/lab_host="XLI-Blue MRF."
Location/Qualifiers

BASE COUNT 144 a 120 c 141 g 172 t 8 others
ORIGIN

Query Match 6.1%; Score 116; DB 22; Length 585;
Best Local Similarity 73.7%; Pred. No. 8.9le-199;
Matches 294; Conservative 0; Mismatches 90; Indels 15; Gaps 13;
Db 50 GCCATGGTCTTTTGAAGGGAGCTATCAGCGCTATCCAGCTCAAGTCACGTCACGTTAAACAG 109
QY 70 GCATGGTCTTACCGTAGCGCTTCGTGCTGCTCAAGCTCCGCTCCGAGCGGTCAA 129
Db 110 CATTCCAGTCTCAGCAATCTCTTAGATGGCTTCAATGCAGTCTCTTCGGATCTTGAT 169
QY 130 CAGTCAATGTTAGCAATCTCTGGCTGGCTTCAAGTCCAAAGCTCTTCGGTCTTGAT 189
Db 170 CTCATTCGCGCTG--A-AGCAATGATCTCTGAACACAGGAGCGCTGAAGAACTG 226
QY 190 CTCGGTCTCGAGCTGTACAGAATGTCTCTGAACACAGGATCGCTGAAAGATC 249
Db 227 AGGCAGAACTGGNAAAGTCAACTTGGGAATATTACTGTGATGGTCATTGGTGA 286
QY 250 AAGTCAGATATG--AAAGTTCAA--TTGGACATCAGATGATGTTGTTGGTGA 306
Db 287 ATGAGAGGATGACAGNGTTCCTTTGGGAAACCTCATTAC--TTG--CCCAGATGGGGAA 344
QY 307 ATGAGAGGAATGACAG--GATTACTGG--AAACCTCATTACCTTGACCTGATGAGGAA 364
Db 345 TTGCTTTTGGGCGCTTGTC--ATTCCGNAGTGTCAAAACACTATTACAGCTGCAAA--CCTG 402
QY 365 TTGCTTTCCGGGGTGTGCTATACCTGATGCCAAAGGTATACCTGACCAAGCCTG 424
Db 403 ATGAGAACCT--TT--CGTNAGGCTCTT--TATGGCTTCTTT 438
QY 425 GGGGTGAGCCCTTGCTGAAGTCTTCTCTGCTCTCTT 463

Search completed: Sat Oct 23 16:24:08 1999
Job time : 3223 secs.

WORLD

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 17:41:52 1999; MasPar time 2810.72 Seconds
1529.581 Million cell updates/sec
Tabular output not generated.

Title: >US-08-702-718-3
Description: (1-1551) from US08702718.seq
Perfect Score: 1551
N.A. Sequence: 1 TCCTCTAACCTTGACCTTCG.....AAAAA.....TTTTTTTTTTTTTT
Comp: AGGAGATTGGAAGTGAAGC.....TTTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_of 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_v1
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_rc
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 11.735; Variance 9.742; scale 1.205

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	1551	100.0	1551	25	A46546 Sequence 2 from Patent	0.00e+00
2	1551	100.0	1551	27	B.vulgaris mRNA for cit	0.00e+00
3	741	47.8	1733	27	CMU19481 Citrus maxima citrate	0.00e+00
4	726	46.8	1747	25	A46547 Sequence 3 from Patent	0.00e+00
5	726	46.8	1747	27	NTCITSYNH N. tabacum mRNA for cit	0.00e+00
6	721	46.5	1859	27	AB017159 Daucus carota mRNA for	0.00e+00
7	664	42.8	1891	27	STCITS S. tuberosum mRNA for m	0.00e+00
8	664	42.8	1891	25	A46345 Sequence 1 from Patent	0.00e+00
9	608	39.2	1680	27	ATMCTITRN Arabidopsis thaliana m	5.28e-285
10	566	36.5	1378	27	PHCITSYNH Populus hybrid mRNA fo	1.94e-263
11	319	20.6	1401	31	AF047042 Homo sapiens citrate s	1.75e-137
12	318	20.5	1455	23	PICGITSYN Porcine citrate syntha	5.60e-137
13	215	13.9	1481	27	SCMTCISNA S. cerevisiae mitochond	2.56e-85

14	215	13.9	2427	27	SCCS01	Yeast gene for citrate	2.56e-85
15	215	13.9	4391	27	SCYNR001C	S. cerevisiae chromosom	2.56e-85
16	215	13.9	23901	27	SCN201952	S. cerevisiae N2019. N2	2.56e-85
17	194	12.5	24643	27	SPAC6C3	S. pombe chromosome 1 c	6.46e-75
18	186	12.0	1680	27	YSCCII2	Yeast (S. cerevisiae) c	5.70e-71
19	186	12.0	5430	27	SCI	Yeast (S. cerevisiae) c	5.70e-71
20	186	12.0	315339	27	SCCHRIII	S. cerevisiae chromosom	5.70e-71
21	172	11.1	1800	27	AB001565	Candida tropicalis DNA	4.33e-64
22	112	7.2	1710	27	ASNCJ11	Aspergillus niger mRNA	3.93e-35
23	101	6.5	562	34	KLAJ9626	Cluyveromyces lactis D	5.81e-30
24	80	5.2	2638	27	NEUCIT1A	N. crassa mitochondrial	2.72e-20
25	81	5.0	48096	21	CET20G5	Caenorhabditis elegans	9.58e-21
26	78	5.0	467	34	GI3212	human STS WI-12459.	2.18e-19
27	73	4.7	118327	28	ATAC004521	Arabidopsis thaliana c	3.81e-17
28	72	4.5	1497	21	TETCJTSYN	T. thermophila mRNA fo	1.06e-16
29	70	4.5	152184	19	CEY70G10	Caenorhabditis elegans	8.20e-16
30	62	4.0	2398	28	ENU89675	Emricella nidulans ci	2.61e-12
31	61	3.9	2790	27	SCCIR3SYN	S. cerevisiae citrate s	7.06e-12
32	61	3.9	17660	27	SCU723	S. cerevisiae chromosom	7.06e-12
33	61	3.9	38779	27	SCU31900	Saccharomyces cerevisi	7.06e-12
34	61	3.9	165536	27	SCCHRXVI	S. cerevisiae chromosom	7.06e-12
35	48	3.1	7218	25	166494	Sequence 14 from paten	2.03e-06
36	38	2.5	826	22	OVU47812	Onchocerca volvulus cy	1.73e-02
37	38	2.5	10772	21	AF012089	Drosophila melanogaste	1.73e-02
38	37	2.4	1797	27	ATFERROCH	Arabidopsis thaliana m	4.12e-02
39	37	2.4	10772	21	AF012089	Drosophila melanogaste	4.12e-02
40	35	2.3	1329	27	ATH6787	Arabidopsis thaliana m	2.26e-01
41	36	2.3	1436	28	AF033539	Lolium perenne caffeic	9.68e-02
42	36	2.3	2286	21	PFCPK	P. falciparum cpk gene.	9.68e-02
43	36	2.3	11829	22	AE001376	Plasmodium falciparum	2.26e-01
44	36	2.3	15892	22	AE001419	Plasmodium falciparum	9.68e-02
45	36	2.3	216021	31	HUAC004787	Homo sapiens Chromosom	9.68e-02

ALIGNMENTS

RESULT	1	A46546	Sequence 2 from Patent WO9524487.	PAT	07-MAR-1997
LOCUS	DEFINITION	1551 bp	DNA		
ACCESSION		A46546			
NID		G2300709			
VERSION		A46546.1	GI:2300709		
KEYWORDS					
SOURCE					
ORGANISM		Beta vulgaris			
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
		euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
		Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.			
REFERENCE		1 (bases 1 to 1551)			
AUTHORS		Mueller-Roeber,B., Landschuetz,V. and La.C.U.			
TITLE		PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN			
JOURNAL		PLANTS			
COMMENT		Patent: WO 9524487-A 2 14-SEP-1995;			
		HOECHST SCHERING AGREVO GMBH (DE)			
		Other publication DE 4438821 960425			
		Other publication DE 4435366 960328			
		Other publication AU 2067995 950925			
		Other publication DE 4408629 950914.			
FEATURES		Location/Qualifiers			
source		1..1551			
		/organism="Beta vulgaris"			
		/strain="ZUCHTLINIE 5S 0026"			
		/db_xref="taxon:3555"			
		/tissue_type="LEAF"			
		/clone="PSBCS"			
		1..1313			
		/EC_number="4.1.3.7"			
		/note="Protein sequence is in conflict with the conceptual			
		translation."			
		/codon_start=1			
		/product="CITRATE SYNTHASE"			
		/protein_id="CAA02909.1"			
		/db_xref="PID:e306423"			

/db_xref="pid:g2300710"

/db_xref="GI:2300710"

/translation="SSNDURSELQELIPEQERLKKIKKEFGSFLGNINVDMLGG
MRGTLGLWETSLIDPEEIRGRFSIPCEOKLLPAAAGAEPLPEGLWLLITGKVP
SKEOVDALSDLRKASIPDHVYKTTIDALPIAHPTQFCTGVMALOTREFOKAYEK
GTHKSKFPEYEDCLSLIAOVVAAVYRMVKNQOVIPLDPSLDYGNFAEWLGF
DSQMLELMRLVTHSHDEGNVSAHGLVGLSPDPLSFAALNGLAGPLHGLA
NOEVLUMIKSVVDEGENISTEQUADYVWKLNSGKVPVGFGLVLRKTDPRYTCQE
FAKHLDPDPFQVLSKLYEVVPIILLKGVKNPVPNDVSHSGLVLRKTDPRYTCQE
TVLFGVSPSISGICSLIWDRLGLPLERPKSVTWLEWLEKFKRPA"

BASE COUNT 455 a 268 c 370 g 458 t

ORIGIN

Query Match 100.0% Score 1551; DB 25; Length 1551;

Best Local Similarity 100.0% Fred No 0.00e+00;

Matches 1551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TCCTCTAACCTTCACCTTCGTTACAGATTACAGAACTGATTCCTGAAACAGCAAGCA 60
Qy 1 TCCTCTAACCTTCACCTTCGTTACAGATTACAGAACTGATTCCTGAAACAGCAAGCA 60
Db 61 CTGAAGAAGATAAAGAAAGAAATTTGGAAGTTTCCAGCTGGGGAATATCAATGTTGACATG 120
Qy 61 CTGAAGAAGATAAAGAAAGAAATTTGGAAGTTTCCAGCTGGGGAATATCAATGTTGACATG 120
Db 121 GTATTGGCGGAATGAGAGGATGACTGCTTTACTTTGGGAGACTTCCTTACTCGACCCA 180
Qy 121 GTATTGGCGGAATGAGAGGATGACTGCTTTACTTTGGGAGACTTCCTTACTCGACCCA 180
Db 181 GAAGAGGATATCCGGTTTCAGGGGTTTTCTATACCTGAAATGCAGAAAGCTTTTACCCGGT 240
Qy 181 GAAGAGGATATCCGGTTTCAGGGGTTTTCTATACCTGAAATGCAGAAAGCTTTTACCCGGT 240
Db 241 GCAAGTCTGGTGCAGAGCCATTGCCTGAAGTCTTCTTTGGCTTCTTTTAAACCGGAAG 300
Qy 241 GCAAGTCTGGTGCAGAGCCATTGCCTGAAGTCTTCTTTGGCTTCTTTTAAACCGGAAG 300
Db 301 GTTCTAGCAAAACAGCAAGTATAGTCTATCAGCAGATTTACGAAAGCTGCTCTATC 360
Qy 301 GTTCTAGCAAAACAGCAAGTATAGTCTATCAGCAGATTTACGAAAGCTGCTCTATC 360
Db 361 CCAGACCATGTGTACAAAACAAATTTGATGCTCTACCTATTACGGCTCATCCATGACTCAG 420
Qy 361 CCAGACCATGTGTACAAAACAAATTTGATGCTCTACCTATTACGGCTCATCCATGACTCAG 420
Db 421 TTTTGCACCTGTTATGSCCTTACAGCTCGAAGCGAATTTTCAAGAGGCATATGAGAA 480
Qy 421 TTTTGCACCTGTTATGSCCTTACAGCTCGAAGCGAATTTTCAAGAGGCATATGAGAA 480
Db 481 GGGATCCATAAGTCAAAAGTTTGGGAGCCCAACATATGAGGACTGCCTTAGTTTGTGCT 540
Qy 481 GGGATCCATAAGTCAAAAGTTTGGGAGCCCAACATATGAGGACTGCCTTAGTTTGTGCT 540
Db 541 CAAGTCTCTGTTGTCAGCTTATGTTATTCGAGAGATGTATAAGAAATGCAAGTAATA 600
Qy 541 CAAGTCTCTGTTGTCAGCTTATGTTATTCGAGAGATGTATAAGAAATGCAAGTAATA 600
Db 601 CCGCTGATCAGCTCCCTTGATTTAGTGGAAATTTCCGACACATGTTGGGATTGATAGC 660
Qy 601 CCGCTGATCAGCTCCCTTGATTTAGTGGAAATTTCCGACACATGTTGGGATTGATAGC 660
Db 661 CCTCAGATGTTGAGCTGATGCGCTTTATGTGCACAAATTCACAGTATCATGAGGGTGA 720
Qy 661 CCTCAGATGTTGAGCTGATGCGCTTTATGTGCACAAATTCACAGTATCATGAGGGTGA 720
Db 721 AATGTTAGTCACACACTGCGCAATTTGGTGGGTAGTCCACATTCAGATCCTTATTGTCA 780
Qy 721 AATGTTAGTCACACACTGCGCAATTTGGTGGGTAGTCCACATTCAGATCCTTATTGTCA 780
Db 781 TTTGCAGCAGCATTAATAGTTTGGCTGGGCCACTCCATGGATTAGCCACAGGAAGTC 840
Qy 781 TTTGCAGCAGCATTAATAGTTTGGCTGGGCCACTCCATGGATTAGCCACAGGAAGTC 840

Db 841 CTGCTGTGGATTAAATCAGTTGTTGATGAATGTGGAGAGAACATCTCGACAGACGAGTTG 900
Qy 841 CTGCTGTGGATTAAATCAGTTGTTGATGAATGTGGAGAGAACATCTCGACAGACGAGTTG 900
Db 901 AAAGATTATGTTTGGAGACACATAACAGTGGCAAGGTTTACCTGGATTGTGCTAGGA 960
Qy 901 AAAGATTATGTTTGGAGACACATAACAGTGGCAAGGTTTACCTGGATTGTGCTAGGA 960
Db 961 GTATTGGCGAGACAGATCCAAAGATACACATGCGCAAGAGAAATTTGCGTTCAAGACATTG 1020
Qy 961 GTATTGGCGAGACAGATCCAAAGATACACATGCGCAAGAGAAATTTGCGTTCAAGACATTG 1020
Db 1021 CCTGATGACCAATTTTCAATTGCTCAAAAGTTGCTATGAAGTGTGCTGCTCTATTCTA 1080
Qy 1021 CCTGATGACCAATTTTCAATTGCTCAAAAGTTGCTATGAAGTGTGCTGCTCTATTCTA 1080
Db 1081 TTAGAGCTTGGAAAGTAAAGAAATCCATGCGCTAAATGTTGATGCTCATAGTGGAGTTTG 1140
Qy 1081 TTAGAGCTTGGAAAGTAAAGAAATCCATGCGCTAAATGTTGATGCTCATAGTGGAGTTTG 1140
Db 1141 CTGAACCACTATGTTTGCAGAAAGCAAGATACATATACGCTTTGTTGGGGTATCAAGG 1200
Qy 1141 CTGAACCACTATGTTTGCAGAAAGCAAGATACATATACGCTTTGTTGGGGTATCAAGG 1200
Db 1201 AGCTTGGAAATATGCTCACAGCTTATATGGACCGAGCTCTTGCGCTGCCCTAGAGAGG 1260
Qy 1201 AGCTTGGAAATATGCTCACAGCTTATATGGACCGAGCTCTTGCGCTGCCCTAGAGAGG 1260
Db 1261 CCNAAGAGTGTCACTATGGAATGGCTTGAAGGTTTGTAAAAAGAGAGCATAAACATTGA 1320
Qy 1261 CCNAAGAGTGTCACTATGGAATGGCTTGAAGGTTTGTAAAAAGAGAGCATAAACATTGA 1320
Db 1321 TGACATATCACTACACTGTTGTTCTTTTGCAGATCTACAATAATATATAGTTGAGGGACAA 1380
Qy 1321 TGACATATCACTACACTGTTGTTCTTTTGCAGATCTACAATAATATATAGTTGAGGGACAA 1380
Db 1381 GAAAGAAATTTATTTTCGGAGATGAGATAAGCGAGGACTCAGAAACATAGTTTCTTTGT 1440
Qy 1381 GAAAGAAATTTATTTTCGGAGATGAGATAAGCGAGGACTCAGAAACATAGTTTCTTTGT 1440
Db 1441 CTCTGTGTGAGTTGGGTTTATATATATTTCACTTTGTAATATATTTGTTGTTTCTTGA 1500
Qy 1441 CTCTGTGTGAGTTGGGTTTATATATATTTCACTTTGTAATATATTTGTTGTTTCTTGA 1500
Db 1501 TCAAAACATGAGATAAGAGTTTTCATAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
Qy 1501 TCAAAACATGAGATAAGAGTTTTCATAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551

RESULT 2 BVCITSYNH 1551 bp mRNA PLN 07-OCT-1996
LOCUS B.vulgaris mRNA for citrate synthase.
DEFINITION X84228
ACCESSION g1556379
NID X84228.1 GI:1556379
VERSION cit1 gene: citrate synthase
KEYWORDS beet.
SOURCE ORGANISM
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;
Caryophyllales: Caryophyllales: Chenopodiaceae: Beta.
REFERENCE 1 (bases 1 to 1551)
AUTHORS La Cognata, U., Landschuetz, V., Willmitzer, L. and Mueller-Roeber, B.
JOURNAL Plant Cell Physiol. In press
REFERENCE 2 (bases 1 to 1551)
AUTHORS La Cognata, U.
JOURNAL Direct Submission
TITLE Submitted (23-JAN-1995) U. La Cognata, Institute fuer
Berlin, FRG Genbiologische Forschung, Berlin GmbH, Ihnestrasse 63, 14195
FEATURES Location/Qualifiers
source 1..1551
/organism="Beta vulgaris"

[illegible]

```

REFERENCE 1 (bases 1 to 1733)
AUTHORS Canel,C., Bailey-Serres,J. and Roose,M.
TITLE Physiological and molecular genetic studies of acid accumulation in
JOURNAL citrus fruits
REFERENCE 2 (bases 1 to 1733)
AUTHORS Canel,C.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-1995) Camilo Canel, University of California at
Riverside, Botany & Plant Sciences, Riverside, CA 92521-0124, USA
FEATURES
Source
1..1733
/organism="Citrus maxima"
/strain="Stamess sweet (pummelo 2240)"
/note="There is a single copy of the cit gene in the
citrus genome. A single transcript with an estimated
length of 1,850 nucleotides can be detected in juice
cells"
/db_xref="taxon:37334"
/sex="hermaphrodite"
/cell_type="juice cell"
/tissue_type="juice tissue"
/dev_stage="immature"
1..56
5'UTR
transit_peptide 57..111
/gene="cit"
/function="targets the protein to the mitochondria"
matrix"
57..1472
/gene="cit"
57..1472
/gene="cit"
/codon_start=1
/function="synthesis of citrate from oxaloacetate and
acetylCoA"
/product="citrate synthase precursor"
/protein_id="AAA82743.1"
/db_xref="PID:g624676"
/db_xref="GI:624676"
/translation="MASLRSAIALSLRBSRACQSNLSNRYWLMQSSADLDLHSQL
KEMIPQERLKKVSDGLGKALGNITIDVIGMGMGTGLLWETSLDDPEGIRFRG
LSIPQCKLLPAAPKDPGLPEGLLWLLTGKVPKQVQDGLSKELDRATVPDYVK
AIDALPVAHPMTOPASGMALQVSEFQEARKEIHKSKSWEPTSDSLNLIARVPV
VAAYVQRIYKDKIIPKDDSLDGNFSLHMLKPLDMLRYVTIHSDBGGNV
SAHTHLVASALSDPLSLAALNGLAGPLHLANOELLWIKSVDPDEGNVTEOL
KDYVWKLNSGVKVPFGHGLRKTDPRYTQREFALKLPDPLFQLVSKLYEYVPP
ILTKLGKVNPNWPNVDHSGVLLNHFGLAEARYITVLFGVSRSLGICSLIWDRLGL
PLERPKSVILDMIEKNCKAA"
111..1469
mat_peptide
/gene="cit"
/function="synthesis of citrate from oxaloacetate and
acetylCoA"
/product="citrate synthase"
1473..1733
polya_signal
1710..1716
polya_site
1733
BASE COUNT 451 a 324 c 446 g 512 t
ORIGIN
Query Match 47.8%; Score 741; DB 27; Length 1733;
Best Local Similarity 78.2%; Pred. No. 0.00e+00;
Matches 1028; Conservative 0; Mismatches 287; Indels 0; Gaps 0;
Db 163 CTGATCTTGACCTTCATCTCAGCTCAAGGAAATGATCCAGAACACAGAGGAGCCCTGA 222
QY 5 CTAACTTGACCTTCGTTTCAGAGTTACAGAACTGATTCCTGAACACAGGACACTGA 54
Db 223 AGAAGTGAAGTCAGACTGTGGAAAACCCAGCTGGGAATATTACCAATTGACGTGGTA 282
QY 65 AGAAGATAAAGAAAGAAATTGGAAAGTTCCAGCTGGGGAATATCAATGTTGACATGGTAT 124
Db 283 TTGGCGGGATGAGAGGAATGACTGGCTTACTTTGGGAACCTCATCTACTTGACCTGATG 342

```

```

QY 125 TGGCGGGAATGAGAGGAATGACTGGTTTACTTTGGGAGACTTCCTTACTCGACCCAGAG 184
Db 343 AGGAATTCGCTTCGGGGCTGTCTATTCTGAGTGCCAGAAAGTTGTTACCTGCGGCTA 402
QY 185 AGGTATCCGGTTCAGGGGTTTCTATCTACTGAATGCCAGAAACTTTTACCCGCTGCA 244
Db 403 AGCCAGATGAGAAACCTTTCGCTGAGGGTCTCTTTGGCTTCCTTTAACTGGAAAGTAC 462
QY 245 GTGCTGTGAGAGGCCAATCCCTGAAGGTCTCTTTGGCTTCCTTTAAACCGGAAGGTC 304
Db 463 CGACCAAGAGCAGGTGCGATGGCTTGTCACAGGAATTTGACAGATCGTGCCACTGTCCAG 522
QY 305 CTAGCAAGAGCAAGTAGATGCTCTATCAGCAGATTTCAGAAACGTCGCTCTATCCAG 364
Db 523 ATTATGTGTAAGGCCATTGATGCTCTACTGTATGGGCACATCGCATGACTCAGTTTG 582
QY 365 ACCATGTGTAACAAACAATTTGATGCTCTACTCTATTACGGCTCATCCAATGACTCAGTTT 424
Db 583 CTAGTGGGGTCATGGCTCTGCAGGTGCAAACTGCAAACTTTCAGGAGGCTTATGAGAAGGAA 642
QY 425 GCATGTGTGTTATGGCTCTTACAGACTCGAAGCGAAATTTACAGAGGCATATGAGAAAGGA 484
Db 643 TTCAATAATCAAACTCATGGGAACCAACATCTGAGGATCTCTTAATTTGATTGCTCGTG 702
QY 485 TCCATAGTCAAGTTTTGGGAGCCACATATGAGGACTGCTTACTGTTGATTGCTCAAG 544
Db 703 TGGCGGTAGTGTGATCATAGTTTATCAAGGATATACAAAGATGGAAGGATATTGCCA 762
QY 545 TTCTGTGTTGTCAGCTTATGTTTATCGGAGGATGTATAAGAATGGAACAAGTAATACCGC 604
Db 763 AGGATGACICACTGGATTATGTTGGGAACCTCTCGGCACATGTTGGGATTCGATCATCTA 822
QY 605 TGAATGACTCCCTGATTTATGGTGGAAATTTCCGACACATGTTGGGATTTGATAGCCCTC 664
Db 823 AGATGCTAGAGCTTATGAGCTTTATGTTACCATTCACAGTACCACAGAGGATGGAATG 882
QY 665 AGATGCTAGAGCTTATGAGCTTTATGTTACCATTCACAGTACCACAGAGGATGGAATG 724
Db 883 TCAGTGTCTACACGGGCCATCTAGTTCGAAGTGCATTTTCAGATCCTTATCTTTCTTCT 942
QY 725 TTAGTGACACACTGGCTCATTTGGTGGTATGATTCATCTTCAGATCTTATTTGCTATTG 784
Db 943 TGGCGCATTAATAATGGTTTACGTGGACACTCCATGGTTTGGCTTAATCAGGAAGTTTGC 1002
QY 785 CAGCAGCATTAATAATGGTTTGGCTGGGCCACTCCATGATTAGCCACAGGAGTCTGCTG 844
Db 1003 TTTGGATCAAAATCTGTGGTGGATGAGTGTGGTGAGAACGTAAACAGGAGCAGTTGAAAG 1062
QY 845 TGTGGATTAATCAATGTTGATGAATGTGGAGAGAACTCTGCACAGAGCAGTTGAAAG 904
Db 1063 ATTAGTCTGGAACCAATTAATGATGGCAAGGTGTGCTCGGATTTGGTCATGTTGTTT 1122
QY 905 ATTAGTCTGGAACCAATTAATGATGGCAAGGTGTGCTCGGATTTGGTCATGTTGTTT 964
Db 1123 TGCCTAAACTGATCCGAGATACACTTGTGAGAGGAGTTTGGCTTTGAAGCATTTGCTG 1182
QY 965 TGGGAAGACAGATCCCAAGATACACATGCCAAGAGAAATTTGCTTGAACCACTTGCCTG 1024
Db 1183 ATGATCAATGTTCCAGCTGGTTTCTAAGCTGTATGAAGTGTGTCCTGCCCCCAATATTACAA 1242
QY 1025 ATGACCAATTTTCAATTTGGTGTCAAAGTTGTATGAAGTGTGCTGCTCTCTATTAG 1084
Db 1243 AACTFAGGCAAGGTTAAACCCCTGGCCAAATGTTGATGCTCACAGTGGGGTCTGCTGA 1302
QY 1085 AGCTTGAAGAGTTAAAGATCCATGGCTAATGTTGATGCTCATAGTGGAGTTTTCGTA 1144
Db 1303 ACCATTTGGTTGGCTGAGGCAAGATATTATCTGTTCTTTTGGAGTATCAAGGAGTC 1362
QY 1145 ACCACTATGGTTTGACAGAAGCAAGATACTATACGGTTTTTGGTTGGGATCAAGGAGTC 1204
Db 1363 TTGGGATCTGCTCTCAGTTTCATATGGACAGCAGCTCTTGGCTGCCACTTTGAGAGGCCAA 1422

```



```

5
RESULT
LOCUS NTCITSYNH 1747 bp mRNA PLN 07-OCT-1996
DEFINITION N.tabacum mRNA for citrate synthase.
ACCESSION X84226
NID g1556428
VERSION X84226.1 GI:1556428
KEYWORDS citl gene; citrate synthase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanales; Solanaceae; Nicotiana.
REFERENCE
1 (bases 1 to 1747)
La Cognata,U., Landschuetz,V., Willmitzer,L. and Mueller-Roeber,B.
Plant Cell Physiol. In press
REFERENCE
2 (bases 1 to 1747)
La Cognata,U.
Direct Submission
Submitted (23-JAN-1995) U. La Cognata, Institute fuer
Genbiologische Forschung, Berlin GmbH, Innestrasse 63, 14195
Berlin, FRG
FEATURES
source
location/Qualifiers
1..1747
/organism="Nicotiana tabacum"
/cultivar="SNN"
/db_xref="taxon:4097"
70..1479
/gene="citl"
70..1479
/gene="citl"
/codon_start=1
/product="citrate synthase"
/protein_id="CAA39008.1"
/db_xref="PID:el37432"
/db_xref="PID:g1556429"
/db_xref="GI:1556429"
/db_xref="SPTREMBL:O24135"
/translation="MVYRGVSLKSLRAVOOTNLNSVRVLOVOTSSGLDRLSEL
QLIEPQQORLKKLSEHGKVGNIIVDMVLGMRGMLGLWEISLSDPDGFRG
LSIEQOQRLVPAKPGGPELGLLWLLTKVPSKEQVLSLSOELRSTVDPHYK
TIDALPVAHPOTFATGMALQVSEFQKAYEKIHKSLWEFTYDSMLIAQVPL
VAAYVRYRMKNGTIPKDDSLDYGANFHLGFSMDHMLKLYVTIHDHGGNV
SAHTGLVASLSDVPLSFPAALGLAGPLHGLANOELLWKSVVECGENISKEOL
KYAWKTLKSGVVPFGHGLRKTDPRTVQOREFALHLPEDLPFLQVAKLYEVFLQ
FLONLAKLPWNPVDAHSGVLLNYYGLTEARYYTVLFGVSRALGICSLINDRALGLP
LERPSYHWEHLENCKKA"
BASE COUNT 490 a 335 c 400 g 522 t
ORIGIN
Query Match 46.8% Score 726; DB 27; Length 1747;
Best Local Similarity 78.5% Pred. No. 0.00e+00;
Matches 1021; Conservative 0; Mismatches 277; Indels 3; Gaps 2;
Db 172 TCTTCTGGTCTTGTATCGCTTCTGAGCTGCAAGAATTGATTCAGAAACACAGATCCG 231
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1 TCCTCAACCTTGACCTTCGTCAGAGTTACAAGAAGTCTGATTCCTGAACAACAGAACGA 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 CTAAAGAGCTCAAGTCAGAGATGGAAGGTTCAATTGGGAACATCACAGTTGATATG 291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 CTGAAGAAGATAAAGAAGAAATTTGAAGTTTCCAGCTGGGGAATCAATGTTGACATG 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 GTTCTTGGTGAATGAGAGGAATGACAGGATTAAGTGGGAACACCTTCATTGACCCG 351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 GTATTGGCGGAATGAGAGGAATGACTGGTTTACTTTGGAGACTTCCTTACTCGACCA 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GATGAAGAAATTCGCTTCGGGCTTGTCTATCTATGAATGCCAAAGATTAATCTGCA 411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 181 GAAGAGGATTCGGGTTACGGGTTTTCTATACCTGAATACCTGAATACCTGCTTTACCGCT 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 GCAAGGCTCGGGGAGAGCCCTTGCCTGAAGGTCCTCTGCTGCTCTTTTAAACAGGAAG 471
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 241 CCAAGTCTGCTGTCAGAGCCATTGCCTGAAGGTCCTCTTTGGCTTCTTTTAAACGGAAAG 300
Db 472 GTGCCATCAAAAGAGCAAGTGGATTTCATTGTCTCAGGAATTCGGAAGTCGTGTACTGTC 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 301 GTTCTAGCAAGAGCAAGTAGATGCTCTATCAGCAGATTTACGAAAACCGTGTCTTCTATC 360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 CCCGATCATGTATACAAAACATTAATGATGCCTTACCAGCTCACAGCTCATCAATGACTCAG 591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 361 CCAGACCATGTGTACAAAACAAATTTGATGCTCTACCTATTACGGCTCATCAATGACTCAG 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 TTTCGTACTGAGTCATGGCTCTTCAGGTTCAAAAGTCAAAATTTCAAAGGCATATGAGAAA 651
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 421 TTTTGCATGTGTATTGGCTTTACAGACTCGAAGCGAATTTCAGAAGGCATATGAGAAA 480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 GGGATTCAAAATCAAAAGTTATGGGAACCGACATATGAGGATTCAGGATTTGATTGCT 711
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 481 GGGATCCATAAGTCAAAAGTTTGGGAGCCACATATGAGGACTGCCCTAGTTGATTGCT 540
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 CAAGTTCACATTTGCTGCTTATGTTTATGCGAGGATGTACAGAAGACGCAACACTATA 771
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 541 CAAGTTCCTGTTGTCAGCTTATGTTTATCGGAGGATGTATAGAATGGACAAGTAATA 600
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 CTAAGGATGACTCAGCTGGATTATGGTGCAAAATTTTCTCACATGCTTGGTTTCAGTAGC 831
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 601 CCGCTGGATGACTCCCTTGATTATGGTGGAAATTTTCGCACACATGTTGGGATTTGATAGC 660
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TCTGACATGATGAGCTTATGAAGCTCTATGTACAGTACACAGTATGATGAAGTGGT 891
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 661 CCTCAGATGCTTGAGCTGATGCGCCTTTATGTCAAAATTCACAGTATGAGGGTGA 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 AACGTGAGTGTCTACAGAGTCACTTGTGCTAGTGTCTGTGACAGACCTTACCTCTCC 951
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 721 AATGTTAGTCACACACTGCCCATTTGGTGGTAGTCCACTTTCAGATCCTTATTGTCA 780
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 TTCGCTGCTGCTTGAATGTTTAGCTGGAGCACCTTCATGCTTTAGCAATCAGGAAGTT 1011
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 781 TTTGAGCAGCATTAATGTTGGTGGGCGCACTTCCATGATGATTAAGCAACAGGAAGTC 840
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1012 TTGCTATGGATCAATCTGTTAGAGAGTGTGGGAGAGACATTTCCAAGAGCAGTTG 1071
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 841 CTGCTGTGATTAATCAGTTGTTGATGATGTGGAGAGAACATCTCGACAGAGCAGTTG 900
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1072 AAAGACTACGCTTGGAAAAACATTTGAAAGTGGCAAGGTTTCCCTGGTTTTCGGACATGGA 1141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 901 AAAGATTATGTTTGGAAAGACACTAAACAGTGGCAAGCTTGTACCTGCATTTGGTCTAGGA 960
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 GTTCTGCGCAAGCTGATCCAGATACACATGCCAGAGAGAGTTCGCTTTGAAGCAATTG 1191
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 961 GTATTGGGGAAGCAGATCCAAAGATACACATGCCAAAAGAGAAATTTGGGTTGAAGCAGTTG 1020
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1192 CCTGAAGATCCACTGTTTCAACTGTTGCAAAACTCTACGAAGTGTTCCTCCAAATCTT 1250
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1021 CCTGATGACCCATTTTCAATTGTTGTCAGAGTTGTATGAGTGGTGCCTCTATTCTA 1080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1251 ACAGAACTTGGCAAGTTAA--ACCTTGGCCAAATGTTGATGCCACAGTGTGTGTTG 1308
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1081 TTAGAGCTTGGAAAGGTAAGAATCCATGSCCTAATGTTGATGCTCATAGTGGAGTTTG 1140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1309 TTGAACATATTGCTTAACTGAAGCAGATATTATACGGTCTCTTTGGTGTATCAAGA 1368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1141 CTGAACCACTATGTTTGCAGAAAGCAGACTATACGGTTTGTGTTGGGTATCAAGG 1200
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1369 GCTCTTGCATTTGCTCTCAGCTAATTTGGGACCGAGCTCTTGGATTGCCACTAGAGAGG 1428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1201 AGTCTTGAATATGCTCACAGCTTATATGGACCGAGCTCTTGGCTTGGCGCTAGAGAGG 1260
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1429 CCAAAGAGTGTCAAAATGGAGTGGCTTGGAGAACCATTTGCCAA 1469
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1261 CCAAAGAGTGTCACTAGGAATGGCTTGAAGAAATTTGTA 1301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 6 AB017159 1859 bp mRNA PLN 01-SEP-1998
LOCUS

```


QY 1264 AAGAGTGTCACTATGGAATGGCTTGAATA 1292

RESULT 7 STCITS 1891 bp mRNA PLN 20-MAR-1996
 LOCUS S.tuberosum mRNA for mitochondrial citrate-synthase.
 DEFINITION X75082
 ACCESSION
 NID 9483509
 VERSION X75082.1 GI:483509
 KEYWORDS mitochondrial citrate synthase.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Asteridae; Solanaceae; Solanales; Solanaceae; Solanum; Potatoe;
 section Petota.
 REFERENCE 1 (bases 1 to 1891)
 AUTHORS Landschuetz V.
 TITLE Direct Submission
 JOURNAL Submitted (20-SEP-1993) V. Landschuetz, Inst. fuer Genbiologische
 Forschung, Berlin GmbH, Ihnestr. 63, 14195 Berlin, FRG
 REFERENCE 2 (bases 1 to 1891)
 AUTHORS Landschuetz, V., Muller-Rober, B. and Willmitzer, L.
 TITLE Mitochondrial citrate synthase from potato: predominant expression
 in mature leaves and young flower buds
 JOURNAL Planta 196 (4), 756-764 (1995)
 MEDLINE 96073431
 FEATURES Location/Qualifiers
 1..1891
 /organism="Solanum tuberosum"
 /cultivar="Desiree"
 /db_xref="taxon:4113"
 /dev_stage="mature"
 /clone_lib="leaf library"
 /clone="5-25-10"
 73..1488
 /EC_number="4.3.1.7"
 /note="ethanolamine ammonia-lyase"
 /codon_start=1
 /product="mitochondrial citrate-synthase"
 /protein_id="CAA52976.1"
 /db_xref="PID:9483510"
 /db_xref="GI:483510"
 /translation="MVFYRSVLSLRSRAVQSNVSRWLOVQSSGLDLRSEL
 VOELIPQOQLKKIKSDMGSGNITVDMVLGMRGNTGLWKPHYLDPDGIRFRG
 LSIPECKVPAKPGCEPLEGLLWLLTGKVPKQVNSIVSGIAESGILSLIMY
 TTIDALVYTHPTQFATGMALQVQSEFQKAYEKIHKSKYWEPTIEDSNMLIAQVP
 LVAIVIRMKNGDITIPKDESLOYGANFAHMLGFSSEMHELLMLRYIYTHSDDEGG
 NVSAHTGHLVASALSDPYLSFAALNGLAGPLHGLANQVLLWIKSVVEECENISKE
 QLDKYVMTKNSKVPFGHGVLRKTVPRYTCQREFAMKHLPEPLFQLVSKLYEVF
 LLFQNLAKLPKPNVDAHSGVLLNYGLTEARYYTVLFVGSRALGICSQLIWDRLG
 LPLRPKSVTMEWLENCKKA"
 BASE COUNT 512 a 370 c 425 g 584 t
 ORIGIN

Query Match 42.8%; Score 664; DB 27; Length 1891;
 Best Local Similarity 78.5%; Pred. No. 0.00e+00;
 Matches 997; Conservative 0; Mismatches 261; Indels 12; Gaps 10;

Db 206 TACAGAATTGATCTCTGAACACAGGATCCCTGAAAAGATCAAGTCAGA-TATGAA 264
 |||||
 QY 29 TACAAGAAGTATCTCTGAACACAGGACACTGAAGAAGATAAAGAAATTTGGAA 88
 |||||
 Db 265 GTTTCAT-TGGG-AACATCAGTGTATGATGTTCTTGTGGTGAATGAGAGGATGACAG 322
 |||||
 QY 89 GTTTCAGCTGGGATATCAATGTTGACATGGTATTGGCGGGAATGAGAGGATGACTG 148
 |||||
 Db 323 GATTACTGTGGAACACTCATACCTTGACCTGATGAGGGAATTCGCTTCGGGGGTGT 382
 |||||
 QY 149 GTTACTTTGGGAGACTTCCTTACTCGACCCAGAGAGGGTATCCGGTTCAGGGGTTTT 208
 |||||
 Db 383 CTATACCTGAATGCCAAAAGGTATTACCTGCGAACAGCCCTGGGGGTGAGCCCTTGCCCTG 442
 |||||

QY 209 CTATACCTGAATGCCAAAACCTTTACCCGCTGCAAGTCTGCTGCGAGACCAATTCGCTG 268
 |||||
 Db 443 AAGGTCTTCTGCTGCTCTTTTAAACAGGAAGGTGCCATCAAAAAGAGCAAGTGAATCAA 502
 |||||
 QY 269 AAGGTCTTCTGCTGCTCTTTTAAACGGAAGGTCTCTAGCAAAAGAGCAAGTAGATGCTC 328
 |||||
 Db 503 TTGCTCAGGAATTCGAGTGGGCATCATATCCCTGATCATCATATACAACTATTG 562
 |||||
 QY 329 TATCAGCAG-ATTACGAAAACGGTCTTC-TATCCCAGACCATG-TGTACAAACAAATG 385
 |||||
 Db 563 ATGCTTTACCACTCACAGCTCATCCCAATGACCAGTTCCTTACTGGAGTCATGGCTCTTC 622
 |||||
 QY 386 ATGCTTACCTTATACGCTCATCCCAATGACTCAGTTTTCCTGCTGTTATGGCCTTAC 445
 |||||
 Db 623 AGTTTCAAGTGAATTTCAAAAGGCATACAGAAAGGATTACAAATCAAGTATTGGG 682
 |||||
 QY 446 AGACTCGAAGCGAATTTTCAAGGCATATAGAAAGGGATCCATAAGTCAAGTTCATGGG 505
 |||||
 Db 683 AACCAACATATGAGGATTCCTGAATCTGATTCCTCAAGTTCCTGCTGCTGCTTATG 742
 |||||
 QY 506 AGCCAACATATGAGGACTGCTTAGTTGATGCTCAAGTTCCTGCTGTTGAGCTTATG 565
 |||||
 Db 743 TTTATCGCAGGATGTACAAGAATGGTGACACTATACCTTAAGGATGAATCCCTGGATTG 802
 |||||
 QY 566 TTTATCGGAGGATGTATAAGAATGGCAAGTAATACCGCTGGATGACTCCCTTGATTATG 625
 |||||
 Db 803 GTGCAAAATTTTGCTCACATGCTGGTTCCTGATCTGCAATGCAATGCAATCTTCATGA 862
 |||||
 QY 626 GTGAAAATTTGCGACACATGTTGGGATTTGATAGCCCTCAGATGCTTGGAGCTG---ATGC 682
 |||||
 Db 863 GGCTCTATGTAAACAATACACAGTATCATGAAGTGGTGAATGTCAGTGTCTACACACGGTC 922
 |||||
 QY 683 GCCTTATGTCACAATTCACAGTATCATGAGGTGGAATGTTAGTGCACACACTGGCC 742
 |||||
 Db 923 ACTTGGTGTAGTGTGTTGTCTGATCCCTTACCTTCCTTTGCTGCTGCTTGAATGGTT 982
 |||||
 QY 743 ATTTGGTGGTAGTCCACTTTTCAGATCTCTTATTTGCAITTTGCGAGCAGCATTAATGSGTT 802
 |||||
 Db 983 TAGCGGGACCACTTCATGGTTTAGCCAATCAGGAAGTTTGTCTATGATGATAAATCTGTTG 1042
 |||||
 QY 803 TGGCTGGCCACTCCATGGATTAGCCAACAGGAAGTCTGCTGTGATTAATCATGTTG 862
 |||||
 Db 1043 TAGAAGATGTGGGAGAACATTTCCAAAGACAGTGTGAAGACTATGTTTGGAAACAT 1102
 |||||
 QY 863 TTGATGAATGTGGAGAGAACATCTCGACAGCAGTGTGAAGATTTATGTTTGAAGACAC 922
 |||||
 Db 1103 TGAACAGTGGCAAGGTGTCCCTGGTTCGGACATGGAGTCTTCGCAAGAGCTGTACCAA 1162
 |||||
 QY 923 TAAACAGTGGCAAGGTGTACCTGGATTTCCTGTAGGAGTATTTCGGAAGACAGATCCAA 982
 |||||
 Db 1163 GATATACATGCCAGAGAGAGTTCCTATGAAGCATTTGCTCTGAAGATCCACTGTTTCAAC 1222
 |||||
 QY 983 GATACACATGCCAAGAGAAATTTCCGTTGAAGCATTGCTGTGATGCCCATTTTTCAT 1042
 |||||
 Db 1223 TGGTTTCAAACTCTACGAAAGTTTTCCTCTGTTCTTACAGAACTTGGCAAGTTAA-A 1280
 |||||
 QY 1043 TGGTGTCAAACTTCTATGAAGTGGTGCCTCTATCTATTAGAGCTTGGAAAGGTAAAG 1102
 |||||
 Db 1281 A-CCTTGGCCAAATGTTGATGCCACAGTGGTGTGTTGTTGAACATTATTGTTTAACTG 1339
 |||||
 QY 1103 ATCCATGGCCCTAAATGTTGATGCTCATAGTGGAGTTTGTGTAACCACTATGTTTTCACAG 1162
 |||||
 Db 1340 AAGCAAGATATTATACGGTCTCTTGGCGTATCAAGAGCTCTTGGCATTTCCTTCAGC 1399
 |||||
 QY 1163 AAGCAAGATATTATACGGTCTTGGTGGTATCAAGGAGTCTTGGCAATATGCTCACAGC 1222
 |||||
 Db 1400 TAATTTGGGACCGAGCTCTTGGATTGGCGGTAGAGAGGCCAAAGAGTGTACAAATGGAGT 1459
 |||||
 QY 1223 TTATATGGGACCGAGCTCTTGGCTTGGCTGCTAGAGAGGCCAAAGAGTGTCACTATGAA 1282
 |||||
 Db 1460 GGCTTGAGAA 1469
 |||||

QY 1283 GGCTTGA AAA 1292

RESULT 8 A46545 1891 bp DNA PAT 07-MAR-1997
 LOCUS
 DEFINITION Sequence 1 from Patent WO9524487.
 ACCESSION A46545
 NID 92300707
 VERSION A46545.1 GI:2300707
 KEYWORDS
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Asteridae; Solanales; Solanaceae; Solanaceae; Solanaceae; Solanaceae;
 section Petota.
 REFERENCE 1 (bases 1 to 1891)
 AUTHORS Mueller-Roeber, B., Landschuetz, V. and La, C.U.
 TITLE PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN
 PLANTS
 JOURNAL Patent: WO 9524487-A 1 14-SEP-1995;
 HOECHST SCHERING AGREVO GMBH (DE)
 COMMENT Other publication DE 4438821 960425
 Other publication DE 4435366 960328
 Other publication AU 2067995 950925
 Other publication DE 4408629 950914.
 FEATURES
 source Location/Qualifiers
 1..1891
 /organism="Solanum tuberosum"
 /strain="C.V. DESIREE"
 /db_xref="taxon:4113"
 /tissue_type="LEAF"
 /clone_lib="CDNA LIBRARY IN PBLUESCRIPTS"
 /clone="PCBS"
 73..1488
 /EC_number="4.1.3.7"
 /codon_start=1
 /product="CITRATE SYNTHASE"
 /protein_id="CAA02908.1"
 /db_xref="PID:e306252"
 /db_xref="PID:g2300708"
 /db_xref="GI:2300708"
 /translation="MVFRSVLSLRSRAVQSNVSNVSRVWQVOTSSGLDLRSEL
 VQELIPEQDLRLKIKIMDKMSIGNITVDMVLGMRGTGLLWPKPHYLDPEGRFRG
 LSIPEQKVLPAAPGPEGLPEGLIWLITLTKVPSKEQVNSIVSGIAESGIIILIMY
 TIDALPYAHPTQFATGVMAQVQSEFQKAYEKGIHKSKYWEPTYEDSNLIAQVP
 LVAAYVRMYKNGDTIPKDESVDYGFANFAMHGFSSSEMHELLMLRYVTHSDHEGG
 NVSAHTGLVASLSDPYLSFAALNGLAGPLHGLANQVLLWIKSVVEGGENISKE
 QLKDYWKTLNSGKVPFGHGVLRKIVPRITCOREFAMKHLPEPDLPLQLVSKLIEVF
 LFLFONLAKLPWPNVDHSGVLLNYGLTEARYITVLFVGSRALGICSLIWDRLG
 LPLRPKSVTMEWLENCKKA"
 BASE COUNT 512 a 370 c 425 g 584 t
 ORIGIN

Query Match 42.8%; Score 664; DB 25; Length 1891;
 Best Local Similarity 78.5%; Pred. No. 0.00e+00;
 Matches 997; Conservative 0; Mismatches 261; Indels 12; Gaps 10;

Db 206 TACAAGATTGATCTCTGAACACAGGATCGCTGAAAAAGATCAAGTCAGA-TAIGAAA 264
 QY 29 TACAAGACTGATCTCTGAACACAGGATCGCTGAAAAAGATCAAGTCAGA-TAIGAAA 88
 Db 265 GGTTCAT-TGGG-ACATCAGATTGATATGGTCTTGGTGGATGAGAGGAATGACAG 322
 QY 89 GTTTCAGCTGGGGAATCAATGTTGACATGGTATGGGCGGAATGAGAGGAATGACTG 148
 Db 323 GATTACTGTGGAACCTCATACCTTGACCTTGATGAGGATTCGCTTCCGGGGTGT 382
 QY 149 GTTACTTTGGGAGACTTCCTTACTCGACCCAGAGAGGGTATCCGGTTCAGGGGTTTT 208
 Db 383 CTATACCTGAATGCCAAGAGGATTAACCTGACGAGAGAGGCTGGGGTGGAGCCCTTGCCTG 442
 QY 209 CTATACCTGAATGCCAAGAGGATTAACCTGACGAGAGGCTGGGGTGGAGCCCTTGCCTG 268

Db 443 AAGGCTCTCTGCTTCTTTTAAACAGAAAGCTGCCATCAAAAGAGCAAGTGAATTCAA 502
 QY 269 AAGGCTCTCTTGGCTTCTTTTAAACCGAAAGGTTCTTAGCAAGAGCAAGTAGATGCTC 328
 Db 503 TTGTCTCAGGAATTCAGAGCTCGGCAATCATATCCCTGATCATCATGTATACAACTATTG 562
 QY 329 TATCAGCAG-ATTACGAAAACGTGCTTC-TATCCAGAGACCATG-TGTACAAAACAATTG 385
 Db 563 ATGCCITACCAGTACAGCTCATCCCAATGACCAGTTTGTCTAGTGGAGTCAATGGCTCTTC 622
 QY 386 ATGCTCTACCTATTACGGCTCATCCAATGACTCAGTTTTCGACTGGTGTATGSCCTTAC 445
 Db 623 AGSTTCAAGTGAATTTCAAAGGCATACGAGAAAGGATTCACAAATCAAGATTAATGGG 682
 QY 446 AGACTCGAAGCAATTTCAAGAGGCATATGGAAGGATCCATAGTCAAGATTTGGG 505
 Db 683 AACCAATATGAGGATTCATGAATCTGATCTCAAGTTCACCTTGTGTGCTTATG 742
 QY 506 AGCAACATATGAGGACTGCTTGTGATTTGATTTGCTCAAGTTCCTGTTGTGCAGCTTATG 565
 Db 743 TTTATCCAGGATTCACAAAGTGGTCACACTATACCTAAGGATGAATCCCTGGATATG 802
 QY 566 TTTATCCAGGATTCATGAAGATGGACAGTAATACCGCTGGATGACTCCCTTGTATG 625
 Db 803 GTGCAAAATTTGCTTCACATGCTTGGTTCAGTCTGAAATGCAATGAATCTTCTTATGA 862
 QY 626 GTGCAAAATTTGCGACACATGTTGGGATTTGATAGCCCTCAGATGCTTGTAGCTG--ATGC 682
 Db 863 GGCTCTATGAACAATACACAGTATCATCAAGGTGGTGAATGTCAGTGTCAACCGGTC 922
 QY 683 GCCTTTATGTACAAATTCACAGTATCATGAGGTGGAATGTTAGTGCACACTGGCC 742
 Db 923 ACTTGTGCTAGTGTCTGCTGATCCTTACCTCTCTCTTCTGCTGCTGCTTCAATGGTT 982
 QY 743 ATTGTGGGTAGTCCACTTTTCAGATCTTATTTGTCATTTGACGAGCAATTAATGGTT 802
 Db 983 TAGCGGACCACTTCATGTTTACGCAATCAGGAAGTTTTCGCTATGATGAATACTGTTG 1042
 QY 803 TGGCTGGCCACTCCATGATGATAGCAACAGGAGTCTCTGCTGTGATTAATCATGTTG 862
 Db 1043 TAGAAGATGTGGGAGAACATTTCCAAAGACAGTGTGAAGACTATGTTTGAAGAACAT 1102
 QY 863 TTGATGAATGTGGAGAGAACATCTCGACAGACAGTGTGAAGATTAATGTTGAAGACAC 922
 Db 1103 TGAACACTGCAAGTGTGCTCCCTGGTTCCTGACATGAGTCTCGCAAGACACTGTACCA 1162
 QY 923 TAAACAGTGGCAGGTTGTACCTGGATTTGGTCTAGAGTATTCGGGAGACAGATCCAA 982
 Db 1163 GATATACATGCCAGAGAGATTCGCTATGAAGCATTTGCCTGAAGATCCACTGTTTCAAC 1222
 QY 983 GATACACATGCCAAGAGAAATTTGCGTTGAAGCACTTGCCTGATGACCCATTTTTTCAAT 1042
 Db 1223 TGGTTTCAAACTCTAGCAAGTTCCTCTGTTCTTACAGAACTTGGCAAGTTAA-A 1280
 QY 1043 TGGTGTCAAAAGTGTATGAAGTGGTGCCTCTATCTATTAGAGCTTGAAGAGTAAAGA 1102
 Db 1281 A-CCTTGGCCAAATGTTGATGCCCCACAGTGGTGTGTTGTTGAACTATTATGTTAACTG 1339
 QY 1103 ATCAATGSCCTAATGTTGATGCTCATAGTGGAGTTTGTGTAACCACTATGTTGACAG 1162
 Db 1340 AAGCAAGATTAATACGGTCCCTCTTTGGCGTATCAAGAGCTCTTGGCATTTGCTCTCAGC 1399
 QY 1163 AAGCAAGATTAATACGGTCCCTCTTTGGCGTATCAAGAGCTCTTGGCATTTGCTCTCAGC 1222
 Db 1400 TAATTTGGGACCCAGCTCTTGGATTCGCGCTAGAGAGCCAAAGAGTGTACACATGAGT 1459
 QY 1223 TTATATGGGACCCAGCTCTTGGCTTGGCTGTAGAGAGCCAAAGAGTGTACATGAAAT 1282
 Db 1460 GGCTTTGAGAA 1469
 QY 1283 GGCTTGA AAA 1292


```

RESULT 9
LOCUS ATMTCTRN 1680 bp mRNA PLN 26-NOV-1997
DEFINITION Arabidopsis thaliana mRNA for mitochondrial citrate synthetase.
ACCESSION X17528
NID 911243
KEYWORDS X17528.1 GI:11243
SOURCE citrate synthase.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1680)
AUTHORS Unger, E.A., Hand, J.M., Cashmore, A.R. and Vasconcelos, A.C.
TITLE Isolation of a cDNA encoding mitochondrial citrate synthase from
Arabidopsis thaliana
JOURNAL Plant Mol. Biol. 13 (4), 411-418 (1989)
MEDLINE 91370823
FEATURES
    source
        1..1680
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
    CDS
        58..1476
        /EC_number="4.1.3.7"
        /codon_start=1
        /product="citrate synthetase"
        /protein_id="CAA35570.1"
        /db_xref="PID:el188578"
        /db_xref="PID:g2652924"
        /db_xref="GI:2652924"
        /translation="MVFFRSVSAFTLRSLRVQGOSSLSNVRWIMOSSTDLDLKSQ
        LQELPEKRLKLIKSEHGKVLQGNITVDVIMVIGMGMTGLLWETSLLDPEEYFALG
        DGLPECKALLPTAQSGSLNHYRFSFVSLNWKGTAKSLKHCRTPWPAAYSDYV
        YNAIDALSTAHPTQPSGYNALQVSEFQKAYENGHIKHFWEPIEDCLNIARV
        PVAAIVYRMTKNGDISPSDKSIDYGANFSLMIGFDDERLKLMLRLTSPSTVMHEG
        NYSATGHLVLSALSDPYLSFAALNGLAGPLHLANOVELLKKSVVEGEGDISKE
        QLKEYVTKLNSGKIPGVYGLVLRNTDPRVVCQREFALKHHPDDPFOCCMLKLAS
        CLTELESEEPNVDHAGKSVLLNHYGLTEARYVYVLFVGSRLSGICSLIWDRELLLA
        LERPSVMDWLEAHCKKASSA"
BASE COUNT 439 a 345 c 408 g 488 t
ORIGIN
Query Match 39.2%; Score 608; DB 27; Length 1680;
Best Local Similarity 77.6%; Pred. No. 5,28e-285;
Matches 1010; Conservative 0; Mismatches 270; Indels 22; Gaps 20;
Db 170 ACCTGACCTGAAGTCGACGTCGACAGAGTTAATCCGGAACACAGAGGACCGCTCTGAAGA 229
QY 8 ACCTTGACCTTCGTTCCAGAGTTACAGAAGACTGATTCCTGAACACAGGAACTGAAGA 67
Db 230 AACTGAAGTCAGAACATGGGAAGGTCCAACTGGGAACATCACTGTTGATATGTTATTG 289
QY 68 AGATAAGAAGAATTTGGAAGTTCCACGCTGGGNAATCAATGTTGACATGGTATTGG 127
Db 290 GTGGAATGAGAGGATGACTGGATGCTTTGGGAAACCTTCATGCTTGACCGGAAGAGG 349
QY 128 GCGGAATGAGAGGAATGACTGGTTTACTTTGGGAGACTTCCTTACTCGACCCAGAAGAGG 187
Db 350 -TATTGCTTTAGGGGATTCGACTTCCTGAGTCGACCAAGACCCCTATTACTACTGCC 408
QY 188 GTATCGGGTTCAGGGGTTTTCTTA-TACCTGAATGCCAGAAA---CTTTACCCGGCTGCA 243
Db 409 CAGTCTGGAGGCTTGAAACCATACCAGA-GGTCTTTTGGCTTCTCTTAACCTGGAAGG 467
QY 244 AGTGCTGGTG-CA-GAGCCATTGCTGAAGGCTCTCTTTGGCTTCTTTAACCGGAAGG 301
Db 468 TACCTTAGCAAGAGCAAGTTGAGCACTGTGCGAAGACTTG-GAAC-CGTGCTGCTGTG 525
QY 302 TTCTT-AGCAAGAGCAAGTAGATGCTCTATCAGCAGATTTACGAAAGAGTCTCTATC 360
Db 526 TCAGATTATGTGTACAAATGCCATGCTGCTGCTCTCCACAGCTCATCAATGACTCAA 585

```

```

QY 361 CCAGACCATGTTGACAAACAATTTGATGCTCTACCTATTACGGCTCATCAATGACTCAG 420
Db 586 TTTGCTAGCGGTGTTATGCGCCCTCCAGGTGCAAGTGAAGTTCAAAAGGCATATGAGAA 645
QY 421 TTTTGCACTGGTGTATGGCTTACAGACTCGAAGCAATTTTCAGAAAGCATATGAGAAA 480
Db 646 GSAATTCATAGTCAAAAGTTCTGGGAGCCCAACATATCAGGATGCCCTCAACCTGATTGCT 705
QY 481 GGGATCCATAGTCAAAAGTTTGGGAGCCCAACATATCAGGACTGCCCTAGTTGATTGCT 540
Db 706 CGTGTCTCTGTTAGCTGCAATGTTTATCGAAGGATGATATAAGAAATGGTGTGATTCATT 765
QY 541 CAAGTCTCTGTTGTTGAGCTTATGTTTATCGGAGGATGATATAAGAAATGGTGTGATTC 600
Db 766 CCCTCAGATAAATCTTTGGATATGTTGGTGCAGAAATTTTCCACATGTTGGGATTTGATG 825
QY 601 CCGTGGATGACTCCCTTGTATGTTGGTGCAGAAATTTCCACACATGTTGGGATTTGATAG 660
Db 826 GAAAGGTTGAAAGAGCTCATGAGGCTT-ACATCACCATCCACAGTATGATCAAGGTGG 884
QY 661 CCTCAGATGCTTGAGCTGATGCGCCTTTATGTCACAATTCACAGTAT-CATGAGGTTG 719
Db 885 AAATGTTAGTGTCTCACACTGGTTCACCTGGTGGTGTGAGTGCACCTTTACAGACCATATC 944
QY 720 AAATGTTAGTGTCTCACACTGGCCTTTGTTGGTGGTGTGAGTGCACCTTTACAGATCCT 779
Db 945 ATTTGCACTGCATTAATGTTAGCTGGCCCACTCCATGTTTGGCTTAATCAGGAAGT 1004
QY 780 ATTTGCACTGCATTAATGTTAGCTGGCCCACTCCATGTTGATAGCCAAACAGGAAGT 839
Db 1005 TTTGCTTTGGATCAATCACTGCTAGAGGATGTGGAGAAGATATATCAAAAGAACAGTT 1064
QY 840 CTTGCTGGATTAATCACTGTTGATGATGTGGAGAGAACATCTCGACAGAGCAGTT 899
Db 1065 GAAAGATATGTTGGAAACATTAACACAGTGGCAAGGTTATTCGGGGATATGCTCAGG 1124
QY 900 GAAAGATATGTTGGAAAGACACTAAACAGTGGCAAGGTTGCTACCTGGATTTGGCTAG 959
Db 1125 TGTCTCGCAATACTCATCAAGATATGTATGCCAAGAGAAATTCGCTTGAAGCATCA 1184
QY 960 AGTATTCGGGAGACAGATCCAGATACACATCCAAAGAGAAATTCGCTTGAAGCACTT 1019
Db 1185 TCCTGACGACCC-TCCTTTCCAGTGTG-CAA-GCT-TATCAAGTTG-GCCTCCTCTCTC 1239
QY 1020 GCCTGATGACCCATTTTTCATTTGGTGTCAAAGTTGTATCAAGTGTGCTCCTATCT 1079
Db 1240 ACT-GAGCTTG-AAAG-TGAAGAACCC-TGGCCAAATGTTGATGCTCAGACTGGGCTTT 1295
QY 1080 ATTAGAGCTTGGAAAGGATAAGAAATCCATGCCCTAATGTTGATGCTCATAGTGGAGTTT 1139
Db 1296 GCTGAACCACTATGCTCTAACCGAAGCAAGGTACTACCCGCTGCTCTTTGGTGTTCAG 1355
QY 1140 GCTGAACCACTATGTTGACAGAGCAAGATACTATACCGTTTGTGGGGTATCAAG 1199
Db 1356 GAGTCTTGGCATGCTCTCAGCTTATATGGACCGAGAGCTCTTGTCTGCATTCAGAG 1415
QY 1200 GAGTCTTGGAAATGCTCAGAGCTTATATGGACCGAGCTCTTGGCTTGGCTGTAGAG 1259
Db 1416 GCCAAGAGTGTACCATCGACTGGCTTGAAGCCCATTTGAA 1457
QY 1260 GCCAAGAGTGTCACTATGAATGGCTTTGAAAAGTTTGTAA 1301
RESULT 10
LOCUS PHCITSYNH 1378 bp mRNA PLN 28-OCT-1996
DEFINITION Populus hybrid mRNA for citrate synthase.
ACCESSION X84227
NID g1556430
KEYWORDS X84227.1 GI:1556430
SOURCE citl gene; citrate synthase.
ORGANISM Populus balsamifera subsp. trichocarpa X Populus deltoides.

```


Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons; Rosidae: Violales: Salicaceae: Populus.

REFERENCE 1 (bases 1 to 1378)
AUTHORS La Cognata,U., Landschuetze,V , Willmitzer,L. and Mueller-Roeber,B.
JOURNAL Plant Cell Physiol. In press
REFERENCE 2 (bases 1 to 1378)
AUTHORS La Cognata,U.
JOURNAL Direct Submission

TITLE Submitted (23-JAN-1995) U. La Cognata, Institute fuer Genbiologische Forschung, Berlin GmbH, Innestrasse 63, 14195 Berlin, FRG
FEATURES
source
1. .1378
Location/Qualifiers
/organism="Populus balsamifera subsp. trichocarpa x Populus deltoides"
/note="hybrid"
/db_xref="taxon:3695"
16. .1035
/gene="citi1"
16. .1035
/gene="citi1"
/EC_number="4.1.3.7"
/codon_start=1
/translation="MPTKEQVQVGLSKELDRALVHDYVFKAIADLPVTHPMTQFATG
VMAIQSEFQKAYEKGPHKSKYWEPTVEDSLIARPIVASYIYRIKDGKVIPI
NDSLIVIGFQSHMLQFSPQMOELRLVVTNHSDEGNVSANTGHVLSALSDPYLS
FAALNGLAGPLHLGLANOELLWIKSVVEEGENITTEOLKDYVWKTLNSGVYVPGF
HGVLKTVPTCYOREFALKHLPDDPLFQVSKLYEVVPPVTLQKVPKPNVDAH
SGVLNITVLTREARYIVLFVSRSGISOLIWDRALGULPLERPKSVTMELLEHNCK
KAPAN"

BASE COUNT 396 a 252 c 313 g 417 t

Query Match 36.5%; Score 566; DB 27; Length 1378;
Best Local Similarity 78.5%; Pred. No. 1.94e-263;
Matches 780; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
Db 14 AGGTACCAACCAAGAGCAAGTCGGTCTTATCGAAGGAATTCGGTCACCGTCTCTGG 73
QY 299 AGGTTCTTAGCAAGAGCAAGTAGATGCTCTATCAGCAGATTTCGAAACGCTGCTTCTA 358
Db 74 TTCATGATTATGCTTTAAGGCCATTGATGCTTACCTGCTCCTCTATCCATGACTC 133
QY 359 TCCAGACCATGTGTACAAACAAATGATGCTCTACCTATTACGGCTCATCCATGACTC 418
Db 134 AGTTTCAACTGCTGTATGCGCCCTACAGGTTTCAGAGTTCAGAGTTCACAAAGCTATGAA 193
QY 419 AGTTTGCAGTGTGTTATGCGCTTACAGCTCGAAGCAATTCAGAGGCAATGAGA 478
Db 194 AAGGATTCATAAATCAAAAGTACTGGAGCCCAACATATGAGGATTCCTCTAGTCTGATTG 253
QY 479 AAGGGATTCATAAAGTCAAAAGTTTGGAGCCCAACATATGAGGACTGCGCTTAGTTGATTG 538
Db 254 CTCGCTGGCAATAGTACTTCATATTTATTCGAAGGATTTACAAAGATGGGAAAGTTA 313
QY 539 CTCAGTTCCTGTGTGAGCTTATGTTATTCGGAGGATGTATAGGAATGACAAAGTAA 598
Db 314 TTCTATGATGACTCTTTGGTTATAGGTGGAAATTTTTCACACATGTTGGATTIGATA 373
QY 599 TACCGCTGATGACTCCCTTGATTGTTGGGAATTTCCACACATGTTGGATTIGATA 658
Db 374 GTCTCGAAATGCAAGAGCTTATGAGCTTTATGTTACTAACCATAGTGACCATGAGGTTG 433
QY 659 GCCTCAGATGCTTGAGCTGATGGCGCTTTATGTGCACAAATTCACAGTATCATGAGGTTG 718

RESULT 11
LOCUS AF047042 1401 bp mRNA PRI 05-JUL-1998
DEFINITION Homo sapiens citrate synthase mRNA, complete cds.
ACCESSION AF047042
NID 93288814
VERSION AF047042.1 GI:3288814
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1401)
AUTHORS Goldenthal,M.J.
TITLE Cloning and sequence analysis of human citrate synthase cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1401)
AUTHORS Goldenthal,M.J.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1998) Molecular Cardiology Institute, 75 Raritan Avenue, Highland Park, NJ 08904, USA
FEATURES
source
1. .1401
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="heart"
/note="isolated from a 22-year old"
1. .1401
CDS

Db 434 GGAATGTCAGTGTCTCATCTACTGGTCACTTGGTTCTAGTGCACCTTCAGATCCTTATCTTT 493
QY 719 GAATGTTAGTGCACACACTGGCCATTTGGTGGTGTAGTCCACTTTTCAGATCCTTATTTGT 778
Db 494 CATTGTCAGTGCATTAATGATGTTTGTAGTGGGCCATCCATGCTTGGCAATCAGGAAG 553
QY 779 CATTGTCAGCAGCATTAAATGTTTGGCTGGGCCACTCCATGATGATAGCAACAGGAAG 838
Db 554 TTCTCTTTGATCAATCGTGTAGAGAGTGTGGAGAGATATAACACACAGACAC 613
QY 839 TCCGCTGTGGATTAATCAGTTTGTATGATGTGGAGAGAACATCTCGACAGACAGT 898
Db 614 TTAAGACTATGTTTGGAAACATTAATAGCGGAAAGTGTGTCTGCAATTTGGTCAIG 673
QY 899 TGAAGATTATGTTTGAAGACACTAAACAGTGGCAAGGTTGTACCTGGATTGGTCTAG 958
Db 674 GAGTTCTACGCAAACTGTACCAAGATATACATGTCAAAGGAGATTTGCATTCAGCAIT 733
QY 959 GAGTATTGCGGAAAGACAGATCCAAAGATACACATSCCAAAGAGAAATTTGCTTGAAGCACT 1018
Db 734 TACTGATGATCCATTGTTCCAGCTGTCTTAAGCTATATGAAGTTGTACCTCCAGTTC 793
QY 1019 TGCTGTATGACCCATTTTTCATTTGTTCAAGTTGTATGAAGTGTGCTTCCCTTATTC 1078
Db 794 TTACCCAGCTGGCAAGGTTTAAACCCCATGGCTTAATGTTGATGCTACAGTGGGGTGT 853
QY 1079 TATTAGAGCTTGGAAAGGTTAAAGAAATCCATGGCTTAATGTTGATGCTATAGTGGAGTTT 1138
Db 854 TGTGAACACTACTAGGTTTAACTGAAGCTAGGTACTACTACTGTTTGGGTGTATCAA 913
QY 1139 TGTGAACACTACTAGGTTTACAGAAAGCAAGATACATACAGGTTTGTGTTGGGTATCAA 1198
Db 914 GGAGTATTGTCATTTGCTCTCAGCTGATATGGATCGAGCTTTGATGTTGCCCTGGAGA 973
QY 1199 GGAGTCTTGAATATGCTCACAGCTTATATGGACCGAGCTCTTGGCTGCCCTAGAGA 1258
Db 974 GGCCAAAGAGTGTACAAATGGAATTCGTTGAGAA 1007
QY 1259 GGCCAAAGAGTGTCACTATGGAATGCTTGAATA 1292

RESULT 11
LOCUS AF047042 1401 bp mRNA PRI 05-JUL-1998
DEFINITION Homo sapiens citrate synthase mRNA, complete cds.
ACCESSION AF047042
NID 93288814
VERSION AF047042.1 GI:3288814
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1401)
AUTHORS Goldenthal,M.J.
TITLE Cloning and sequence analysis of human citrate synthase cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1401)
AUTHORS Goldenthal,M.J.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1998) Molecular Cardiology Institute, 75 Raritan Avenue, Highland Park, NJ 08904, USA
FEATURES
source
1. .1401
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="heart"
/note="isolated from a 22-year old"
1. .1401
CDS

Best Local Similarity 64.6%; Pred. No. 5.60e-137;
Matches 758; Conservative 0; Mismatches 410; Indels 5; Gaps 4;

```

Db 185 TGGGCCAAATCTGTTGGACATGATGTTGGGCGATGAGGCGATGAGGGATTGGTCT 244
QY 98 TGGGGAATCAATGTTGACATGTTGTTGGGCGAATGAGAGGAATGACTGGTTACTTT 157
Db 245 ATGAACATCGGTTCTTGATCTGATGAGGCGATCGGTTTCGAGGCTACATATCCCTG 304
QY 158 GGGAGACTTCCTTACTCGACCCAGAGAGGTTATCGGTTTCAGGGGTTTTTCTATACCTG 217
Db 305 AATGCCAGAAATGCTGCCAAGGCTAAGGGTGGGGAAGACCGCTGCCAGAGGGCTTAT 364
QY 218 AATGCCAGAAATGCTGCCAAGGCTAAGGGTGGGGAAGACCGCTGCCAGAGGGCTTAT 277
Db 365 TTTGGTGTCTGTTAACTGGACAGATCCCAACAGAGGAGGAGTGTCTTGGCTCTCAAAAG 424
QY 278 TTTGGTCTTTTAAACGGGNAAGGTTCTTACGAGGAGGAGGAGTGTCTTGGCTCTCAAAAG 337
Db 425 AGTGGCAAAAGAGGCGAGCTCTCCCTTCCATGTGTCACCATGCTGGACAACTTTCCCA 484
QY 338 ATTTAGAAAACGTCTTCTATCCAGACCATGTGTACAAAACAATTTGATGCTCTACCTA 397
Db 485 CGAATCTACACCCCATGTCTCAGCTCAGTCAGCCCATTTACAGCCCTCAACAGTGAAGTA 544
QY 398 TTACGGCTCATCAATGACTCAGTTTGGCACTGGTGTATGGCCTTACAGACTCGAAGCG 457
Db 545 ACTTTGCCGAGCATATGAGAGGGTATCCACCGAACCAAGTACTGGGAGTTGATTATG 604
QY 458 AATTTCAGAGGCGATATGAGAAAGGATCCATTAAGTCAAGTTTGGGAGCCCAACATAG 517
Db 605 AAGACTGTATGATCTGATCGAAAGCTACCTTGTCTGCAGCAAGATCTACCGGAATC 664
QY 518 AGGACTGCCCTTGTGTTGATGCTCAAGTTCTCTGTTGTCAGCTTATGTTTTCGAGGA 577
Db 665 TCTACCGGGAAGCAGCAGTATTTGGGGCCATGATTTCTAACTGGACTGGTCCCAAT 724
QY 578 TGATAGAATGG-ACAAGTAATACCGC--TGGATGACTCCCTTGAATGATGTTGGAAT 534
Db 725 TCACCAACATGTTAGGCTATACGATGCTCAGTTTCAGGAGTCAAGGCTTGTACCTCA 784
QY 635 TCGCACACATGTTGGGATTTGATAGCCCTCAGATGCTGAGCTGATGCGCCTTATGCA 694
Db 785 CCATCCACATGACCATGAAGGTGGCAATGTAAGTGTCTATACAGCCATTTGGTGGCA 844
QY 695 CAATTACAGCTGATCATGAGGTGGAATGTTAGTGCACACACTGGCCATTTGGTGGTA 754
Db 845 GTGCCCTTTTCAGACCCCTACTTCTCTTTTGCAGCAGCCATGATGGCTGGCAGGCC 904
QY 755 GTCCACTTTTCAGATCCTTATTTGTCATTTTGCAGCAGCATTAATGTTTGGCTGGCC 814
Db 905 TACATGGCTTGGCAAAATCAGAAAGTGTGTTTGGCTGACA-CAGCTGCAGAGGAAGTC 963
QY 815 TCCATGATGATGACCAACAGGAGTCTCTGCTGGATTAATCAGTTGTGATGATGTG 874
Db 964 GCGAAGATGTCAGATGAGAGTTACGAGACTACATCTGGAATACACTCAACTCAGGA 1023
QY 875 GAGAGA-ACATCTCGACAGAGCTTGAAGATTATGTTTGAAGACACTAAACAGTGGC 933
Db 1024 CGGTTGTCCCGCTATGCCCAGCGAGTACTAAGAAAGCATGATCCAGATATACCTGT 1083
QY 934 AGGTTGTACTGGATTGCTAGGAGTATTCGGAAGCAGATCCCAAGATACATGTC 993
Db 1084 CAAGAGAGTTTGTCTGAAACACCTCGCTCATGAGCCCATGTTTAAAGTGGTTGCTCAG 1143
QY 994 CAAGAGATTTGGTTGAGACATTCCTGCTGATGACCCATTTTCAATTTGTTGTCAG 1053
Db 1144 CTGTACAGATTTGGCCCAATGCTCTCGTGAACAGAGGCAAGGCTAAGAAATCCTTGGGCC 1203
QY 1054 TTGTATGAAGTGGTGCCTCTATCTATTAGAGCTTTGGAAGGTAAGAATCCATGGCCT 1113
Db 1204 AATGTGATGCTCACAGTGGGTTGCTCCAGTACTATGGCATGACGAGATGACTAC 1263

```

```

QY 1114 AATGTTGATGCTCATAGTGGAGTTTGTGTAACCACTATGTTTGACAGAGAAGATAC 1173
Db 1264 TACACAGTCTGTTTCGGGTATACACGGGCTGATGAGGCTGATGACACAGCTCATCTGAGC 1323
QY 1174 TATAGGTTTGTGTTGGGTATCAAGGAGTCTTGGAAATATGCTCACAGCTTATATGGAC 1233
Db 1324 CGAGCCTTAGGCTTCCCTCTAGAGAGGCCCAAG 1356
QY 1234 CGAGCTCTGCTTGGCGCTAGAGAGGCCCAAG 1266

RESULT 13
LOCUS SCWTCISNA 1481 bp DNA PLN
DEFINITION S.cerevisiae mitochondrial citrate synthase gene, complete CDS.
ACCESSION Z23259
NID G313749
VERSION Z23259.1 GI:313749
KEYWORDS mitochondrial citrate synthase.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 1481)
AUTHORS Lindner, P.
JOURNAL Thesis (1993) Protein Engineering, Max-Planck-Institut fuer Biochemie
REFERENCE 2 (bases 1 to 1481)
AUTHORS Lindner, P. and Plueckthun, A.
TITLE The effect of point mutations in the hinge of yeast citrate synthase
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1481)
AUTHORS Lindner, P.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1993) Peter Lindner, Protein Engineering, Max-Planck-Institut fuer, Biochemie, Am Klopferspitz 18a, Martinsried, D-82152, Fed. Rep., Germany
REFERENCE 4 (bases 1 to 1481)
AUTHORS Suissa, M., Suda, K. and Schatz, G.
TITLE Isolation of the nuclear yeast genes for citrate synthase and fifteen other mitochondrial proteins by a new screening method
JOURNAL EMBO J. 3 (8), 1773-1781 (1984)
MEDLINE 85003587
REMARK (sites)
FEATURES
source
1..1481
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/strain="D273-10B"
/db_xref="taxon:4932"
/clone="Plasmid pFCS1 (Suissa et al., EMBO J. 3, 1773-1781 (1984))."
1..1481
/gene="CIT1"
/db_xref="SGD:L0000341"
1..1481
/gene="CIT1"
/db_xref="SGD:L0000341"
1..1481
/gene="CIT1"
/db_xref="SGD:L0000341"
/product="mitochondrial citrate synthase"
/protein_id="CA380781.1"
/db_xref="PID:g313750"
/db_xref="GI:313750"
/db_xref="SWISS-PROT:P00890"
/translation="MSAIISTTSKSLRSRSTQCNMOKALFALLNARHYSASBOT
LKPERAEIIPAKAEIKPKKHEKTVIGVLLLEQAVGGMRTGKGLVWEGSVLDPEEG
IRPRTPIEIQRELPAKSGSTPEALFWLLTGTETPDACVKAALSADLAARSHIP
EHVIQLDLSLPKDLHPMAOFSTAVTALESESEKFAKAYAGCVSKKWSYTFEDSLDL"
sig_peptide
CDS
1..1440
/codon_start=1
/db_xref="SGD:L0000341"
/product="mitochondrial citrate synthase"
/protein_id="CA380781.1"
/db_xref="PID:g313750"
/db_xref="GI:313750"
/db_xref="SWISS-PROT:P00890"
/translation="MSAIISTTSKSLRSRSTQCNMOKALFALLNARHYSASBOT
LKPERAEIIPAKAEIKPKKHEKTVIGVLLLEQAVGGMRTGKGLVWEGSVLDPEEG
IRPRTPIEIQRELPAKSGSTPEALFWLLTGTETPDACVKAALSADLAARSHIP
EHVIQLDLSLPKDLHPMAOFSTAVTALESESEKFAKAYAGCVSKKWSYTFEDSLDL"

```

GKLPVIAKIIYNNVFKDKITSDPNADYGNKLAQLLGVENKDFIDLMRLYLTIHSDH
EGGNVSATHTHLVGSALSSPYLSLAAGLHGRANQEVLEWLFKLREVKSDY
SKETLEKYLWDTLNAGRPVPGYCHAVLRKTDPRYTAQREFAKHFDYELFKLVSTIY
EVAPGLVKHKTKNPNVDSHSGVLLQYIGLUTASFYTLVFGVARAIGVLPOLLID
RAVGAPIERPKSFSTERYKELVKIESKN"
172

conflict

/gene="CITI1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace="g"
234^235

conflict

/gene="CITI1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace="a"
1472

conflict

/gene="CITI1"
/citation=[4]
/db_xref="SGD:L0000341"
/replace="a"

BASE COUNT 452 a 266 c 319 g 444 t
ORIGIN

Query Match 13.9%; Score 215; DB 27; Length 1481;

Best Local Similarity 59.4%; Pred. No. 2.56e-85;

Matches 677; Conservative 0; Mismatches 462; Indels 0; Gaps 0;

Db 244 GGTGGTATGAGAGGTATTAAAGCCCTTGTGGGAAGCTTCGGGTATAGACCCCGAAGAA 303
Qy 127 GCGGGAATGAGGAATGACTGTTTACTTTGGGAGACTTCCTTACTCGACCCGAGAAG 186
Db 304 GGTATTAGATTTAGGGTTCGTACTATTCCAGAAATTTCAAAGGAACACTACCAAGGCTGAG 363
Qy 187 GGTATCCGTTACAGGGTTTTCTATACCTGAATCCAGAAACITTTACCCGCTGCAAGT 246
Db 364 GGTAGTACAGAACCTTTGCCAGAACCTTTATTTGGTTCCTTTTGGTCTTTGACTGGTGAATACCT 423
Qy 247 GCTGGTGCAGAGCAATGCTCTGAAGGCTCTCTTTGGCTTCTTTTAAACCGAAAGGTTCTCT 306
Db 424 ACTGAGCGTCAAGTTAAAGCCCTTCTGCTGATTAGTGCACATCAGAAATTTCCAGAG 483
Qy 307 AGCAAGAGCAAGTAGACTGCTCTATCAGCAGATTTACGAAACGTGCTTCTATCCAGAC 366
Db 484 CAGCTTATCCAACTTTTAGATAGCCTCCCAAGATCTACATCCAAATGGCCAAATTTCT 543
Qy 367 CATGTGTACAAAACAATTCATGCTCTACCTATTACGGCTCATCCAATGACTCAGTTTTC 426
Db 544 ATTGCGGTGACTGCTTTAGAAAGCGAGTCTAGTTTGGCAAGCATATGCTCAGGTGTA 603
Qy 427 ACTGGTGTATGGCTTACAGACTCGAAGCGAATTTTCAAGAGGCATATGAGAAAGGATC 486
Db 604 TCCAAGAAAGATATTGGAGCTATACATTTGAAGATTTCGTTAGATCTCTGGTAAATTA 663
Qy 487 CATAGTCAAGTTTGGGAGCCACATATGAGGACTGCCCTTAGTTGATTGCTCAAGTT 546
Db 664 CTGTTATTGCTTCCAAAATTTATCGTAAATGTTTCAAGGATGTTAAATTTACTTCAACC 723
Qy 547 CTGTTTGTGTCAGCTTATGTTTATCGGAGGATGTATAAGATGGCAAGTAAATACCGCTG 606
Db 724 GATCCTTAATGCTGACTATGTAATAATTTGGCCCACTTTTGGGCTAGCAAAAAGAGAT 783
Qy 607 GATGACTCCCTTGATTATGTTGGTGAATTTTCGCACACATGTTGGGATTTGATAGCCCTCAG 666
Db 784 TTTATTGACTTAATGAGACTATATTAACTATTCATTCGTATCATGAAGGTGGTAAAGCTT 843
Qy 667 ATGCTTGAAGTATGAGCTGCTTTATGTCACAAATTCACATGATCATGAGGGTGGAAATGT 726
Db 844 TCTGCCCATACTACAAATTTAGTGGTCTGCTTATCTTCGCCATACTATCTTTGGCC 903
Qy 727 AGTGACACACTGGCCATTTGGTGGTGTAGTCCACTTTCAGATCTCTTATTTGCTCAITTTGA 786
Db 904 GCTGGTTGAATGGTTTACCTGGCCCAATTTACATGCTGCTGCAATCAAGAAGTTTTAGAA 963

Qy 787 GCAGCATTAATAGTGTGGCTGGCCACTCCATGGATTAGCCAAACAGGAAGTCTCTGCTG 846
Db 964 TGGCTATTAAATTCAGAGAGAAGTGAAGAGTCACTATTCAAAGAAGAAACAATTAAGAG 1023
Qy 847 TGGATTAAATCAGTTGTGATGAATGTGGAGAGAACAATCTCGACAGAGCAGTTTGAAGAT 906
Db 1024 TACTTGTGGGACTTTTGAACGCGAGGAGAGTGTCTCTGGTATATGCGCATGGGTTTTG 1083
Qy 907 TATGTTTGAAGACACTAAACAGTGGCAAGTTGTACTCTGGAATTTGGCTAGAGATTG 966
Db 1084 AGAAAACCTGATCCTTCGTTATACGGCTCAACGTGAATTCGCATTTGAACATTTTCCAGAT 1143
Qy 967 CGGAAGACAGATCCCAAGATACACATGCCAAGAGAATTTGCGTTGAAGCACTTTGCGCTGAT 1026
Db 1144 TAGGAGTTATTAAAGTTGGTCTCCACCAATTTATGAAGTTGCCCCAGGGGTTTAACTAAG 1203
Qy 1027 GACCAATTTTTCAATTTGGTGTCAAAGTTGTATGAAGTGTGCGCTCTCTATTATTAGAG 1086
Db 1204 CATGTTAAACCTTAAGAACCCATGGCCAAATGTTGATTCACATTCGGGTGTTTTATTGCAA 1263
Qy 1087 CTGGAAGGTAAGAATCCATGCCTAATGTTGATGCTCATAGTGGAGTTTCTCTGAAC 1146
Db 1264 TACTATGCTTAACGTAGGCTTCGTTCTACACTGATTTGTTGGTGTGGCAGAGCTATT 1323
Qy 1147 CACTATGTTTACAGAGAAGCAAGATACATACGTTTGTGTTGGGTTATCAAGGAGTCTT 1206
Db 1324 GCTGTGTTACCCCAATTAATCATGATAGGCTGTTGGTCTCAATCGAAGGCCAAA 1382
Qy 1207 GGAATATGCTCAGACTTATATGGACCGAGCTCTTGGTTGCGCTAGAGAGGCCAAA 1265

RESULT 14

LOCUS SCDS01 2427 bp DNA PLN 12-SEP-1993
DEFINITION Yeast gene for citrate synthase.
ACCESSION X00782
NID 93602
VERSION X00782.1 GI:3602
KEYWORDS citrate synthase; synthase.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2427)
AUTHORS Suissa, M., Suda, K. and Schatz, G.
TITLE Isolation of the nuclear yeast genes for citrate synthase and
fifteen other mitochondrial proteins by a new screening method
JOURNAL EMBO J. 3 (8), 1773-1781 (1984)
MEDLINE 85003587
COMMENT Data kindly reviewed (09-MAY-1985) by G. Schatz.
FEATURES
Location/Qualifiers
1..2427
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
376..1818
/note="citrate synthase"
/codon_start=1
/protein_id="CAA25359.1"
/db_xref="pid:g3603"
/db_xref="GI:3603"
/db_xref="SWISS-PROT:P00890"
/translation="MSALSTTSKSLRSGTSRQCNMOKALFALLNARHYSASEQT
LKERFAEIIPAQAEIKFKKHKGTGIVLELEQAYGMRGKGLVWGSVLDPPEE
GIRFRGTIPEIQRELFPKAESTEPLEALFWLLTCEIPTDAQKALSADLAARSEI
PEHVIOQLDSLDPKDLHPMAQFSTAVTALESSEKFAKAYAGVSKKWEYTFEDSLD
LGKLPVIAKIIYNNVFKDKITSDPNADYGNKLAQLLGVENKDFIDLMRLYLTIHSD
HEGNVSATHTHLVGSALSSPYLSLAAGLHGRANQEVLEWLFKLREVKGD
YKSETIENKTLHVLGNSALSSPYLSLAAGLHGRANQEVLEWLFKLREVKSTI
YEVAPGLVKHKTKNPNVDSHSGVLLQYIGLUTASFYTLVFGVARAIGVLPOLLID
RAVGAPIERPKSFSTERYKELVKIESKN"

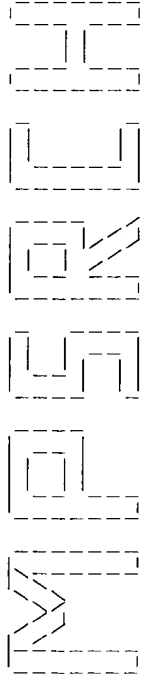
BASE COUNT

757 a 439 c 474 g 757 t

Query Match

13.9%; Score 215; DB 27; Length 2427;

Best Local Similarity 59.4%; Pred. No. 2.56e-85; Matches 677; Conservative 0; Mismatches 462; Indels 0; Gaps 0;									
Db	622	GGTGTATCAGAGGTATTAAGGCGCTTTTGGGAAGGTTCCGTGTGTAGACCCGGAAGAA	681						
Qy	127	GCGGAATCAGAGGAATGACTGGTTTACTTTGGGAGACTTCCCTTACTCGACCAGAGAG	186						
Db	682	GGTATTAGATTAGGGTCTACTATTCCAGAAATTCMAAGGGAATACCAAGGCTGAG	741						
Qy	187	GGTATCCGGTTCAGGGGTTTCTATACCTGAATGCCAGAACTTTTCCCGCTGCAAGT	246						
Db	742	GGTAGTACAGAACCTTTCGCAAGCTTTATTTGGTTGCTTTTGAAGTGGTGAATACCT	801						
Qy	247	GCTGGTGCAGACCAATGCCTGAAGTCTCTCTTGGCTTCITTAACCGGAAGGTTCCT	306						
Db	802	ATGACGCTCAAGTTAAAGCGCTTTCTGCTGATTAGTGCACGATCAGAAATCCAGAG	861						
Qy	307	AGCAAAGACAAAGTAGATGCTCTATCAGCAGATTACGAAACGTGCTTCTATCCAGAC	366						
Db	862	CACGTTATCCAACTTTTAGATAGCTCCCAAGATCTACATCCAAATGGCGAAATTTCT	921						
Qy	367	CATGTGTACAAACAATGTATGCTCTACTCTATTAGCGCTCATCAATGACTCAGTTTGC	426						
Db	922	ATTGCCGTGACTGCTTTAGAAAGCGAGTCTAAGTTTGCCTCAAGCATATGCTCAAGTGA	981						
Qy	427	ACTGGTGTATGCGCTTACAGACTCGAAGCGAATTTTCAGAAGGCATATGAGAAAGGATC	486						
Db	982	TCAGAAAGAATATGAGCTATACATTTCAAGATTCGTTAGATCTGCTGGGTAAATTA	1041						
Qy	487	CATAAGTCAAGTTTGGGAGCAACATATGAGGACTGCCCTAGTTTGTATGCTCAAGTT	546						
Db	1042	CTGTATTGCTTCCAAAATTTATCGTAATGTGTCAAGGATGTAATTAATCTTCAACC	1101						
Qy	547	CTGTGTGTGCAGCTTATCTTTATCGAGGATGTATAGATGACAAAGTAATACCGCTG	606						
Db	1102	GATCCTATGCTGACTATGTGTAATAATTTGGCCCACTTTTGGCTAGCAAAACAAGAT	1161						
Qy	607	GATGACTCCCTTGATATGTGGTGAATTTCCGACACATGTTGGGATTTGATAGCCCTCAG	666						
Db	1162	TTATTGACTTAATGAGACTATATTAACTATTCTATTCATGATCAAGGTGGTAACGTT	1221						
Qy	667	ATGCTTGACTGATGCGCTTTATGTCACATTCACATGATCATGAGGTGGAAATGTT	726						
Db	1222	TCGCGCATACACATATTAGTGGGTCTCGCTTATCTTCGCGCATCTTAICTTTGGCC	1281						
Qy	727	AGTGCACACACTGGCCATTTGGTGGGTAGTCACCTTTTCAGATCCTTATTTCATTTGCA	786						
Db	1282	GCTGTTTGAATGGTTTACTGCGCCCATACATGCTCGTCCCATCAAGAGTTTAAAG	1341						
Qy	787	GCAGCAATTAATGGTTTGGCTGGGCCATCCCAATGATTAGCCACACAGAGTCTCTG	846						
Db	1342	TGCTATTTAAATTTGAGAGAAGAGTGAAGGTGACTATTCAAAGAAACAATTTGAAAG	1401						
Qy	847	TGAATTAATACGTTGTTGATCAATGTGAGAGACATCTCGACAGACAGCTTGAAGAT	906						
Db	1402	TACTTGGGATCTTTGAACCGAGGAGATGTTTCTCGTTATGCGCATCGCGTTTGG	1461						
Qy	907	TATGTTTGAAGACACATAAACAGTGGCAAGTGTACCTGGATTTGGTCTAGGAGTATG	966						
Db	1462	AGAAAACATGATCCTGTTATACGGCTCAACGTGAATTCGATTTGAAACATTTCCACAG	1521						
Qy	967	CGAAGACAGATCCAAGATACACATGCCAAGAGAAATTTGGTTGAACACATTCGCTGAT	1026						
Db	1522	TACGAGTATTTAAGTTGGTCTCCACCATTTTATGAAGTTGCCCGAGGGTTTAACTAAG	1581						
Qy	1027	GACCCATTTTTCAATTTGGTCAAAGTTGTATGAAGTGGTGGCTTCCTATTCTATTAG	1086						
Db	1582	CATGGTAAACTAAGAACCATGGCCAAATGTGATTCACATTCGGTGTTTTATTGCAA	1641						
Qy	1087	CTTGGAAAGGTAAAGAAATCCATGGCCATAATGTGATGCTCATAGTGGAGTTTGTCTAAC	1146						
Db	1642	TACTATGGTCTAACTGAGGCTTCGTTCTACACTGATTTGTTGGTGTTCGACAGACTATT	1701						



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 19:14:04 1999; MasPar time 351.57 Seconds
Tabular output not generated. 945.291 Million cell updates/sec

Title: >US-08-702-718-3
Description: (1-1551) from US08702718.seq
Perfect Score: 1551
N.A. Sequence: 1 TCCTCTAACCTTGACCTTCG.....AAATAAAAAAAAAAAAAA 1551
Comp: AGGAGATTGGAACCTGGAAGC.....TTTTTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.510; Variance 7.769; scale 1.224

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1551	100.0	1551	16	T04200 Sugar beet citrate sy	0.00e+00
2	726	46.8	1747	15	T04201 Tobacco citrate synth	0.00e+00
3	664	42.8	1891	17	T03410 Potato citrate synth	0.00e+00
4	664	42.8	1891	16	T04199 Potato citrate synth	0.00e+00
5	42	2.7	204	1	N81164 Base substituted E.co	3.09e-06
6	42	2.7	204	1	N81164 Base substituted E.co	3.09e-06
7	39	2.5	91	9	Q51746 Oligonucleotide probe	5.63e-05
8	38	2.5	91	9	Q51746 Oligonucleotide probe	1.46e-04
9	35	2.3	1077	46	V39085 Human tumour necrosis	2.46e-03

10	35	2.3	1164	49	V07654	Nucleotide sequence o	2.46e-03
11	32	2.1	114	12	Q7C468	Generic DNA sequence	3.84e-02
12	32	2.1	114	12	Q7C468	Generic DNA sequence	3.84e-02
13	33	2.1	525	46	V41453	Nucleotide sequence o	1.55e-02
14	32	2.1	654	29	T47675	Human G protein gamma	3.84e-02
15	32	2.1	940	28	T58281	Arabidopsis STZ polyp	3.84e-02
16	32	2.1	1505	47	V40522	Homo sapiens C1542.2	3.84e-02
17	32	2.1	1694	50	V33190	Secreted protein B018	3.84e-02
18	32	2.1	2116	6	Q35987	Tomato hsp80 cDNA seq	3.84e-02
19	31	2.0	67	24	T14322	Primer used in the la	9.41e-02
20	31	2.0	208	40	V09116	3' nucleotide sequenc	9.41e-02
21	31	2.0	775	47	V44861	Clone BV66_1 coding s	9.41e-02
22	31	2.0	1046	2	Q10377	Plasmid pMG3C9 used s	9.41e-02
23	31	2.0	1092	38	T91592	Rat UBCE2A coding seq	9.41e-02
24	31	2.0	1136	2	N70435	Sequence encoding ins	9.41e-02
25	31	2.0	1137	30	T63234	Human ubiquitin conj	9.41e-02
26	31	2.0	1558	22	T28255	Survival motor neuron	9.41e-02
27	31	2.0	1560	22	T19828	Human survival motor	9.41e-02
28	31	2.0	1582	22	T18831	Human survival motor	9.41e-02
29	31	2.0	1582	22	T28259	Survival motor neuron	9.41e-02
30	31	2.0	1856	30	T63233	Human ubiquitin conj	9.41e-02
31	31	2.0	2589	30	T61590	Human c-IAP1	9.41e-02
32	31	2.0	3784	1	Q05326	Sequence encoding fis	9.41e-02
33	31	2.0	3850	1	N81634	Sequence of new plasm	9.41e-02
34	31	2.0	3851	33	T90019	Yellow tail/fin tuna	9.41e-02
35	31	2.0	3851	33	T79857	Yellowtail tuna DNA e	9.41e-02
36	31	2.0	3855	43	V22382	Human telomerase reve	9.41e-02
37	31	2.0	4017	43	V22428	Human telomerase reve	9.41e-02
38	31	2.0	5173	36	T89783	Human Immunodeficien	9.41e-02
39	30	1.9	91	46	V44650	Mammalian DNA replica	2.28e-01
40	30	1.9	114	12	Q70468	Generic DNA sequence	2.28e-01
41	30	1.9	524	58	V73005	Human adult testis se	2.28e-01
42	30	1.9	608	5	Q31796	PRP3 gene (cDNA)	2.28e-01
43	30	1.9	684	13	Q72956	Clone pSP65-Xai-11 en	2.28e-01
44	30	1.9	1892	16	Q74063	The mouse beta-actin	2.28e-01
45	30	1.9	3899	51	V63192	cDNA from clone dt674	2.28e-01

ALIGNMENTS

RESULT 1
ID T04200 standard; cDNA to mRNA; 1551 BP.
AC T04200.
DT 25-JAN-1996 (first entry)
DE Sugar beet citrate synthase cDNA.
KW Citrate synthase; flower formation; ss.
OS Beta vulgaris strain Zuchtlinie 5S 0026
FH Key Location/Qualifiers
FT Cds 1..1313
FT WO9524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR P-PSDB: R82839.
DT DNA encoding plant citrate synthase - used to regulate flower formation.
PT To improve storage of tubers, etc. and to reduce sprouting
PS Claim 5; Page 57-60; 87pp; English.
CC To identify a cDNA from sugar beet which codes for citrate
CC synthase, a cDNA bank of leaf tissue from sugar beet was prep.
CC syntheses of this cDNA bank were screened using radioactive DNA
CC probes which comprise a mixture of Solanum tuberosum citrate
CC synthase cDNA (T04199) and Nicotian tabacum citrate synthase cDNA
CC (see T04201). One of the clones was sequenced. The nt. sequence is
CC given in T04200.
SQ Sequence 1551 BP; 455 A; 268 C; 370 G; 458 T;

Query Match 100.0% Score 1551; DB 16; Length 1551;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 1551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tctctaaactgaccttgccttcagagttacagaactgattcctcgaacaacaggaacga 60
 Qy 1 TCGCTAACTTGAUCTTCTGCTCAGAGTTACAAGAACTGATTCCTGGAACAACAGGAACGA 60

Db 61 ctgaagaataaagaagaataatttgaagtttccagctcgggggaatatacaatttcaatg 120
 Qy 61 CTGAAGAATAAAGAAGAATAATTGGAAGTTTCCAGCTGGGGAATATCAATGTTGACATG 120

Db 121 gtattgggggaatgagagaatgactggtttacttttggagacttccctactcaccaca 180
 Qy 121 GTATTGGGGGAATGAGAGAAATGACTGGTTTACTTTGGGAGACTTCCCTTACTCGACCCA 180

Db 181 gaagagggtatccgggttcagggtttttctatataccttgatccgaaactttttaccgcgt 240
 Qy 181 GAAGAGGGTATCCGGTTCAGGGTTCCTTCTATACCTGAATGCCAAGAACTTTTACCOCGT 240

Db 241 gaaagtctggtgcagagccattgctgaaggtctcttcttgggtctcttttaaccgggaag 300
 Qy 241 GCAAGTCTGCTGTCAGAGCCATTGGCTGAAGGTCTCTTGGCTTCCTTTTAAACCGGAAG 300

Db 301 gttcctagcaagagcaagtagatgctctatcagcagatttaccgaaacgtgctctatc 360
 Qy 301 GTTCTTAGCAAGAGCAAGTAGATGCTCTATCAGCAGATTTACGNAACGTGCTTCTATC 360

Db 361 ccagaccatgttacaacaataatgatgctctacattatagcgtctatccaatgaactcag 420
 Qy 361 CCAGACCATGTGTACAAAACAATGATGCTCTACCTATTACGGCTCATCCCAATGACTCAG 420

Db 421 ttgtcactgggtgtatgaccttccagactcgaagcgaatttcagaagcgcatacgaaga 480
 Qy 421 TTTTGCACGTGGTGTATGGCTTTCAGACTCGAAGCGAATTTTCAAGAGGCATATGAGAA 480

Db 481 gggatccataaagtcgaagttttgggagccacaacatataggagctgcttagttgattgct 540
 Qy 481 GGGATCCATAAGTCAAGTTTGGGAGCCACAATATAGGAGCTGCCTTAGTTGATTGCT 540

Db 541 caagttccctgtgttcagacttatgtttatcggagagatgataagaatggacaagtaata 600
 Qy 541 CAAGTTCCTGTGTTGCAGCTTATGTTTATCGGAGGATGTAAAGATGGCAAGTAATA 600

Db 601 ccgctggatgactcccttattgatgtgggaatttcgcacacatgttgggatttgcagc 660
 Qy 601 CCGCTGGATGACTCCCTTGATTATGTTGGTGAATTTTCGCACACATGTTGGGATTGATAGC 660

Db 661 cctcagatccttgagctgtagccttttatgtcacaattcagtgatcatgaggg-gga 720
 Qy 661 CCTCAGATCCTTGAGCTGATGCTGCTTTTATGTCACAATTCACAGTGTATCATGAGGGTGA 720

Db 721 aatgttagtgacacactggccttgggtgggttagtcacactttcagatccctatttctca 780
 Qy 721 AATGTTAGTGCACACACTGGCCATTGGTGGGTAGTCCACCTTTCAGATCCTTATTGTGCA 780

Db 781 ttgcagcagcattaaatgggttggctgggcccactccatggattagccaaccaggaagtc 840
 Qy 781 TTTGCAGCAGCATTAATGGTTGGCTGGGCCACTCCATGGATTAGCCAACCAAGGAAGTC 840

Db 841 ctgctgtggaataaactcagttgtgatgaatgtgagagaacatctcagacagacagttg 900
 Qy 841 CTGCTGTGGAATAAATCAGTTGTGATGAATGTGAGAGAACATCTGCACAGACAGATTG 900

Db 901 aaagattatgtttgggaagacataacagtgagcaggttgcacctggatttggcttagga 960
 Qy 901 AAGATTATGTTTGGGAAGACATAACAGTGGCAGAGTGTACCTGGGATTTGGTCTAGGA 960

Db 961 gtattgggaagacataacagatacacatgcccagaagaatttgccttgaagcaacttg 1020
 Qy 961 GTATTGGGAAGACATGCAAGATACACATGCCAAGAGAATTTGCGTTTGAAGCACTTG 1020

Db 1021 cctgatgaccatttttttcaattgggtcacaagtgtgtaagtggtgtgcctccatttcta 1080

Qy 1021 CCTGATGACCATTTTTCATTTGGTGTCAAGTTGTATGAAGTGTGCTCTCTATTCTA 1080
 Db 1081 tttagagcttggaaagtaaaagatccatggcctaattgtgatgcataatgaatgtttg 1140
 Qy 1081 TTAGAGCTTGGAAAGGTAAAGAAATCCATGGCTTAATGTTGATGTCATAGTGTGAGTTTG 1140

Db 1141 ctgaaccactatggtttgaagaagaagataactatcaggttttgggtgggtatcaagg 1200
 Qy 1141 CTGAACCACTATGTTTGCACAAGCAAGATACTATACGGTTTGTGTTGGGTATCAAGG 1200

Db 1201 agtcttgaataatgctcacagcttatatgggaccgagcctcttggcttgcctcagaagq 1260
 Qy 1201 AGTCTTGGNAATATGCTCACAGCTTATATGGGACCGAGCTCTTGGCTTCCGCTAGAGAGG 1260

Db 1261 ccaagagtgctcaactatggaaatggcttgaagttttgaaagaagaagacataaacttga 1320
 Qy 1261 CCAAGAGTGTCACTATGGAATGGCTTGAAAGTTTTGTAAAGAGAGCATAACTTGA 1320

Db 1321 tgacatatcaactcactcttcttcttcttgcgaatctacaataatagtttggaggacaa 1380
 Qy 1321 TGACATATCAACTCACTCTGTTGTTGTCGAATCTCAATAATATAGTTTGGAGGACAA 1380

Db 1381 gaagaattttatttctcgagatgagataaagcggagcctcagaacaacatagtttcttctg 1440
 Qy 1381 GAAGAATTTTATTTCGAGATGAGATAAGCGAGGACTCAGAAACATAGTTTCTTTGT 1440

Db 1441 ctcttgcaggttgcgttttctatatatttcaacttataatattgtatggtttcttga 1500
 Qy 1441 CTCCTTGCAGTTCGCTTTTATATATTTCACCTTGTAAATATATTGTTGTTCTTGA 1500

Db 1501 tcaaacatgagataaagagtttctataaaaaaataaaaaaataaaaaa 1551
 Qy 1501 TCAAAACATGAGATAAAGAGTTTCTATAAAAAAATAAAAAAATAAAAAA 1551

RESULT 2

ID T04201 standard; cDNA to mRNA; 1747 BP.
 AC T04201;
 DT 25-JAN-1996 (first entry)
 DE Tobacco citrate synthase cDNA.
 KW Citrate synthase; flower formation; ss.
 OS Nicotiana tabacum
 FH Key Location/Qualifiers
 cds 70..1476
 FT /*tag- a
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR P-PSDB; R82840.
 PT DNA encoding plant citrate synthase - used to regulate flower formation.
 PT To improve storage of tubers, etc. and to reduce sprouting
 PS Claim 7; Page 60-63; 87pp; English.
 CC To identify a cDNA from tobacco which codes for citrate
 CC synthase, a cDNA bank of leaf tissue from tobacco was prepd.
 CC Plaques of this cDNA bank were screened using a radioactive DNA
 CC probe which comprises Solanum tuberosum citrate synthase cDNA
 CC (T04199). One of the clones was sequenced. The nt. sequence is
 CC given in T04201.
 SQ Sequence 1747 BP; 490 A; 335 C; 400 G; 522 T;

Query Match 46.8%, Score 726; DB 16; Length 1747.
 Best Local Similarity 78.5%; Pred. No. 0.00e+00;
 Matches 1021; Conservative 0; Mismatches 277; Indels 3; Gaps 2;

Db 172 tcttctgcttgccttgcgttctgagctgcagaattgattccagacaacagatcgc 231

1 TCCTCTAACCTTGACCTTCGTTCCAGAGTTTACAAGAACTGATTCTCTGAACAACAGCAAGCA 60
 Db 232 ctaaaagactcaagtcagagcatgaaaggttcaattgggaaacatcacagtttatatg 291
 Qy 61 CTGAAGAAGATAAAGAAAGAAATTTGAAGTTTCCAGCTGGGAATATCAATTTGACATG 120
 Db 292 gtcttggttggaatgagagaaatgacaggtattactgtgggaacacctcattactaacccc 351
 Qy 121 GTATTGGCGGAATGAGAGGAATGACTGGTTTACITTTGGAGACTTCCTTTACTCGACCCA 180
 Db 352 gatgaaggaattcgcttttcggggcttgctcttatctatgaaatgcaaaaggtatatactgca 411
 Qy 181 GAAGAGGGTATCCGGTTCAGGGGTTTTTCTATACCTGTAATGCGAGAAACTTTTACCCGCT 240
 Db 412 gcaagcctggggagagcccttgctgaggtcttctctgcttcttctttaaagaaag 471
 Qy 241 GCAAGTGTCTGTCAGAGCAATTCCTCGAAGGTCITCTTTTGGCTCTCTTTTAAACCGGAAG 300
 Db 472 gtccatcaaaaagcaagtggtattctatgtctcaggaatttcgaagtcgtgctactgtgc 531
 Qy 301 GTTCTTAGCAAAAGCAAGTAGATGCTCTATCAGCAGATTTACGAAACGCTCTCTATC 360
 Db 532 ccgcatcatgtatataaaactatgtacttaccagtcacagtcacagtcacatgactcag 591
 Qy 361 CCAGACCAATGTACAAAACAATTTGATGCTCTACCTATTACGGCTCATCAATGACTCAG 420
 Db 592 ttgctactgagtcagtcgctctcaggttccaaagtccaagtgaatttcacaaagccatagaaa 651
 Qy 421 TTTTGNCTGGTGTATGGCTTACAGACTCGAAGCGAATTTACAGAGGCATATGAGAAA 480
 Db 652 gggattcaaaaatcaagttatgggaaccgacatagaggattccatgagtttgatgtc 711
 Qy 481 GGGATCCATAAGTCAAAGTTTGGGAGCCAACTATGAGGACTGCTTGTAGTTGATGCT 540
 Db 712 caagttccacttgtctgcttattgtattatcgagagatgataaagaacggaacactata 771
 Qy 541 CAAGTTCCTGTTGTGAGCTTATGTTATGCGAGGATGTATAAGAAATGAGCAAGTAATA 600
 Db 772 cctaagtgactcaactggattatggtgcaaaattttgtctcacaactgttggtttccagtaac 831
 Qy 601 CCCTGAGTACACTCCCTTGATTATGTTGGGAAATTTCCGACACATGTTGGGATTTGATAGC 660
 Db 832 tctgacatgatagcttatgaagctctatgtcacaatcacacagtgatcatcaaggtgtc 891
 Qy 661 CCTCAGATGCTTGAGCTGATGCGCTTTATGTCACAAATTCACAGTATCATGAGGCTGA 720
 Db 892 aacgtcagtcacacagtcactgtgtgtgctagtgctttgtcagacccttaccctcc 951
 Qy 721 AATGTTAGTGCACACTGCGCAATTTGGTGGGTAGTCCACTTTCAGATCCTTATTGTCA 780
 Db 952 ttcgctgctgtttgaaatggttttagctggaccacttcattcgttttagccaatcagaagtt 1011
 Qy 781 TTTGACGACGATTAATATGTTGGTGGGCCACTCCATGATTAGCCCAACGAGGAAGTC 840
 Db 1012 ttgctatggataaacttgttttagagagagtggtgggagaaacatttccaaagagcagttg 1071
 Qy 841 CTCGTGGSATTAATACAGTTGTGTGTAATGTGGAGAGAAATCTCGACAGACAGTGTG 900
 Db 1072 aaagactacgcttggaataacatigaaagtggcaaggtgtgtccctggttctcgagacatga 1131
 Qy 901 AAAGATTATGTTGGAAGACACTAAACAGTGGCAAGGTTGTACTGGATTGGTCTAGGA 960
 Db 1132 gttctgcgaagactgatccaagatacacatgccagagagagtgctcttgaagcatttg 1191
 Qy 961 GTATTGGGAAGACAGATCCAAGATACACATGCCAAAGAGAAATTTGGTGTGAAGCACTTG 1020
 Db 1192 cctgaactcactgtttcactggttgcaaaactctagaagtggtt - cctcctaatttt 1250
 Qy 1021 CCTGATGACCCATTTTTTCAATTGGGTCAAAGTTGTATGATGGTGTGCTCTCTATTTCA 1080
 Db 1251 acgaacttggaagaaagttaa - acccttggccaaagtgttatgccacagtggtgtgtg 1308

Qy 1081 TTAGAGCTTGAAAGGTAAAGAAATCCATGCTTAATGTTGATGCTCATAGTGAGTTTG 1140
 Db 1309 ttgaactattatgatttaactgaagcaagattattatacagctctcttggatgtatcaaga 1368
 Qy 1141 CTGAACCACTATGTTTGACAGAAGCAAGTACTATACGGTTTGTGTTGGGTATCAAGG 1200
 Db 1369 gctcttgacattgctcctcagcttaatttgggaaccgagctcttggattggcactagagag 1428
 Qy 1201 AGTCTTGGAAATATGCTCACAGCTTATATGGACCGAGCTCTTGCTTGCGCTAGAGAGG 1260
 Db 1429 ccaagagtgctcaaatggagtggtggcttgagaacaccattgcaa 1459
 Qy 1261 CCAAGAGAGTGTCACTATGGAATGGCTTGAAGAGTTTGTIAA 1301

RESULT 3

ID T03410 standard: cDNA; 1891 BP.
 AC T03410;
 DT 24-APR-1996 (first entry)
 DE Potato citrate synthase gene.
 KW Citrate synthase; Inhibitor; increased storage capacity; potato;
 KW antisense DNA; ss.
 OS Solanum tuberosum.
 FH Key Location/Qualifiers
 FT cds 73..1488
 FT /*tag= a
 FT /product= potato_citrate_synthase
 PN DE4408629-A1.
 PD 14-SEP-1995.
 PF 09-MAR-1994; 408629.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U; Landschutze V, Muller-Roeber B, Landschutze V;
 DR WPI: 95-321536/42.
 DR P-PDB; R86393.
 FT Inhibiting citrate synthase (CS) activity in plants - to inhibit
 PT flower formation and improve storage capacity, e.g. in potatoes,
 PT also new CS sense and anti-sense DNA sequences
 PS Claim 6; Page 15-19; 35pp; German.
 CC The potato citrate synthase (CS) gene or cDNA sequence encoding CS
 CC (T03410) may be used to produce antisense CS sequences. CS DNA
 CC sequences are useful for altering CS activity in plants. Antisense
 CC CS sequences can be used to inhibit CS expression in plants and has
 CC the effect of inhibiting flower formation and by doing so improves
 CC the plant's storage capacity. This is partic. useful in crop plants
 CC of any kind but esp. useful in potatoes. In addition to altering CS
 CC activity the DNA sequences can also be used to identify similar
 CC sequences in the genomes of other plants and in the production of
 CC transgenic plants with altered CS activity..
 SQ Sequence 1891 BP; 512 A; 370 C; 425 G; 584 T;

Query Match 42.8%; Score 664; DB 17; Length 1891;
 Best Local Similarity 78.5%; Pred. No. 0.00e+00;
 Matches 997; Conservative 0; Mismatches 261; Indels 12; Gaps 10;

Db 206 tacaagaattattctctgaacaacagagatcgctgaaaaagatcaagtcaga-tatgaaa 264
 Qy 29 TACAAGAAGTCTGATCTCTGAACAACAGGAAACGACTGAAGAGATAAAGAAATTTGGAA 88
 Db 265 gtttcaat-tggg-aacatcacagttgatattggttcttggtaagtgaagaaatgacag 322
 Qy 89 GTTTCACAGCTGGGAATATCAATGTTGACATGGTATTGGGCGGGAATGAGGAATGACTG 148
 Db 323 gattactgtgaaacctcattacccttgaccctgatgaggaattcgcttccgggggtgtg 382
 Qy 149 GTTTACTTTGGAGACTTCCTTACTCGACCCAGAGAGGGTATCCGGTTACAGGGTTTT 208
 Db 383 ctataacctgaatgcaaaaggtattacctgcagcaaacctcgggggtgagcccttgctg 442
 Qy 209 CTATACCTGAATGCCAGAAACTTTTATCCCGTGCAGAGTGCCTGCTGCGAGCCATTGCTG 268

Db 443 aaggtctctctgtctctctttaaagaggaagtgccatcaaaagagcaagtgaaattcaa 502
 QY 269 AAGGTCTCTTGGCTCTCTTAAACGGGAAGTTCCTAGCAAGAGCAAGTAGATGCTC 328
 Db 503 ttgtctcagaattgcagagtgccatcatccctgcatcatctgtatatacaattg 562
 QY 329 TATCAGCAG-ATTTACGAAACGCTGCTTC-TATCCAGACCATG-TGTACAAAACAATTG 385
 Db 563 atgcttaccagtcacagctcatccaatgaccagttgtctactgagtcagtcgtcttc 622
 QY 386 ATGCTCTACTATATACGGCTATCCCAATGACTCAGTTTTTGCACCTGGTGTATGGCCTTAC 445
 Db 623 aggttcaagtgaaatttcaaaagcctacacagaaagggattccacaatacaaatattggg 682
 QY 446 AGACTCGAAGCAATTCAGAGGCATATGAGAAAGGGATCCATTAAGTCAAAAGTTTGGG 505
 Db 683 aaccacatataggattccatgaatctctgctcgaagttcccaattgttgcgttatg 742
 QY 506 AGCAACATATGAGGACTGCTTAGTTTGAATGCTCAAGTTCCTGTTGTCAGCTTATG 565
 Db 743 ttatcgcagagtgatcaagaatggtgacactatacctaagaatgaatccctggtattg 802
 QY 566 TTTATCGGAGATGATAGAAATGGACAAGTAATACCGCTGGATGACTCCCTTGATATG 625
 Db 803 gtgcaaatttgtctcacatgctgtgtttcagtagctctgaaatgcataactcttataga 862
 QY 626 GTGGAATTTGACACATATGTTGGGATTTGATAGCCCTCAGATGTTGAGCTG---ATGC 582
 Db 863 ggtctatgaataacacagtcagtcagtgaggggtgtaattctcagtcacacagctc 922
 QY 683 GCCTTTATGCAATTCACATGATCATGAGGGTGAATTTAGTGCACACACTGGCC 742
 Db 923 actgtgtcagtgcttctgtctccttaccctctctctctctctctctctctctctct 982
 QY 743 ATTGTGGGTAGTCACATTTTCAGATCCTTATTTGTCTATTTGCAGCAGCATTAATGTT 802
 Db 983 tagccggaccactctatggttttagccaatcaggaagttgtgtatgataaaatctgtg 1042
 QY 803 TGCTGGGCACCTCCATGATGATTAGCAACACAGCAAGTCTGCTGTGGATTAATCAGTTG 862
 Db 1043 tagaagaattggggagaacatttccaaagagcagttgaaagactatgtttgaaacat 1102
 QY 863 TTGATGAATTTGGAGAGACATCTCGACAGACAGTGAAGATTTATGTTTGGAGAGAC 922
 Db 1103 tgacagtggaagtggtctccctggttttgacatggagttctgcgaagactgtaccac 1162
 QY 923 TAAACAGTGCAGAGTTGTACCTGGATTGGTCTAGGAGTATTGCGGAAGACAGATCCAA 982
 Db 1163 gatatacgcagagagaggttgcgtatgaacatttgctgcaagatccactgtttcaac 1222
 QY 983 GATACATGCCAAGAGAAATTTGCGTTGAAGCATTGCTGATGACCCATTTTTCAT 1042
 Db 1223 tggtttcaaaactctacgaagttt-cctccgtgtcttcaagaacttggcaagttaa-a 1280
 QY 1043 TGTGTCAAAGTTGTATGAAGTGGTCCCTCTATTTCTATTAGCTTGGAAAGGTAAAGA 1102
 Db 1281 a-ccttggccaaatgttgatgccacagtggtgtgtgttgtaactattatggtttacgt 1339
 QY 1103 ATCCATGGCCTAATGTGTATGCTCATAGTGGAGTTTGTGTAACCACTATGTTTGCAG 1162
 Db 1340 aaccaagatatatacagtcctcttggcgatcaagagctcttggcatttgcctcagc 1399
 QY 1163 AACCAAGATACTATACGGTTTTTGTGTGGGGTATCAAGGAGTCTTGGAAATATGCTCACAGC 1222
 Db 1400 taatttggaccagctcttggattgcccgtcagagagggccaaagagtgtcacaaatggagt 1459
 QY 1223 TTATATGGGACCGAGCTCTTGCTTGGCTTGGCCGTAGAGAGGCCAAGAGTGTCACTATGGAT 1282
 Db 1460 gctttgagaa 1469
 QY 1283 GGCTTGAATA 1292

RESULT 4
 ID T04199 standard; cDNA to mRNA; 1891 BP.
 AC T04199; 1996 (first entry)
 DT 25-JAN-1996
 DE Potato citrate synthase cDNA.
 KW Citrate synthase; flower formation; tuber storage; ss.
 OS Solanum tuberosum.
 FH Key Location/Qualifiers
 FT CDS 73..1485
 PN WO9524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI: 95-328278/42.
 DR P-PSDB: R82838.
 PT DNA encoding plant citrate synthase - used to regulate flower formation.
 PT To improve storage of tubers, etc. and to reduce sprouting
 PS Claim 5; Page 53-56; 87pp: English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of
 CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. A cDNA library was prepd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 1891 BP; 512 A; 370 C; 425 G; 584 T;

Query Match 42.8%; Score 664; DB 16; Length 1891;

Best Local Similarity 78.5%; Pred. No. 0.00e+00;

Matches 997; Conservative 0; Mismatches 261; Indels 12; Gaps 10;

Db 206 tacaagaattgattcctgaacaacagagatcgctgaaaaagatcaagtcaca-tatgaaa 264
 QY 29 TACAAAGACTGATTCCTGAAACACAGGAGCACTGAAGAAGATAAAGAAATTTGGAA 88
 Db 265 ggttcaat-tggg-aacatcacagttgatatggttctgtggaatgagaggaatgacag 322
 QY 89 GTTTCAGCTGGGGAATATCAATGTTGACATGTTATTGGCGCAATGAGAGGAATGACTG 148
 Db 323 gattactgtggaacacctcattaccctgacctgatgaggaattcgctccgggggttgt 382
 QY 149 GTTTACTTTGGGAGACTTCCTTACTCGACCCAGAGGGGTATCCGGTTCAGGGGTTTT 208
 Db 383 ctatacctgaatgcacaaagggtattaccctgcagcaagcctggggtgagcccttgccgt 442
 QY 209 CTATACCTGAAATGCCAAGAACTTTTACCCGCTGCAAGTGTGTGCGAGACCATTCGCTG 268
 Db 443 aaggtctctctggctctcttttaacaggaagaagtgccatcaaaagagcaagtaattcaa 502
 QY 269 AAGGTCTCTTGGCTCTCTTTTAAACCGGAAAGGTTCCTAGCAAGAGCAAGTAGATGCTC 328
 Db 503 ttgtctcagaattgcagagtcggggaatcctcctgatcatcatcatatatacaactattg 562
 QY 329 TATCAGCAG-ATTTACGAAACAGTGTCTT-TATCCCAAGACCATG-TGTACAAAACAATTG 385
 Db 563 atgcttaccagtcacagctcatcccaatgaccagttgttgcactgagtcagtcgtcttc 622
 QY 386 ATGCTCTACCTATATACGGCTATCCCAATGACTCAGTTTTCGACTGGTGTATGGCCTTAC 445
 Db 623 aggttcaaatgtaatttcaaaagggatcacagaaagggattccacaatacaaatattggg 682
 QY 446 AGACTCGAAGCAATTTTCAAGAGGCATATGAGAAAGGGATCCCATTAAGTCAAAAGTTTGGG 505
 Db 683 aaccaacatatgagagttccatgaatctgtctcagttcactgctgcttattg 742

	Query Match	2.7%	Score 42;	DB 1;	Length 204;
	Best Local Similarity	12.2%	Pred. No. 3.09e-06;		
	Matches	14;	Conservative	60;	Mismatches 40;
				Indels	1;
				Gaps	
Db	76	ydcchvgcgymrttthhyrmrbnvdydnrsdaaawccyrrsvkydcycnaxchddh	135		
QY	778	TCATTGTGCAGCAGCATTA-AATGGTTTGGCTGGCGCACTCCATGATGATTAGGCCAACAGGA	836		
Db	136	yyvbbbyvnhnnnccbnhvhcnvbnhrnwayvyrhdarrddvhecvc	190		
QY	837	AGTCCTGCTGTGGATTAAATCAGTTGTTGTAAGTGTGGAGAGACATCTCGACA	891		

RESULT	6	
ID	N81164 standard; DNA; 204 BP.	
AC	N81164;	
DT	08-NOV-1990 (first entry)	
DE	Base substituted E.coli beta-galactosidase alpha-fragment.	
KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.	
OS	Escherichia coli.	
Key	Location/Qualifiers	
FT	misc_feature 19..69	
FT	/*tag= a	
FT	/function= multiple cloning site	
FT	187..204	
FT	/*tag= b	
EP	285123- A.	
PN	05-MAY-1988.	
PD	30-MAR-1988; 105163.	
PF	(SUPO-1987; US-034819.	
PR	(3-APR) SUOMEN SOKERI OY.	
PI	Lehtovaara P, Knowles J, Koivula A, Reinikainen T;	
PR	WPI: 88-279927/40.	
PT	Introducing random point mutations into nucleic acids -	
PT	by prep of single stranded template, annealing a primer, elongation,	
PT	misincorporation, completion of molecules and screening.	
PS	Disclosure: p: English.	
CC	Random point mutations were introduced into the alpha fragment of	
CC	E.coli beta-galactosidase. The wild type sequence was obtained as a	
CC	single stranded template and an oligonucleotide was hybridised to	
CC	it to generate a popn of DNA molecules which terminate at all	
CC	possible nucleotide positions within a specified region. The	
CC	variable 3' ends generated in this way are used as primers for	
CC	reverse transcriptase. Nucleotides are misincorporated by the	
CC	transcriptase and the molecules are completed to forms that can be	
CC	amplified and then expressed in a suitable host-vector system.	
CC	The sequence covers all 176 diffit base substitutions, most of which	
CC	occurred singularly in any given mutant.	
CC	See also P80575.	
SQ	Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;	
Query Match	2.7%; Score 42; DB 1; Length 204;	
Best Local Similarity	10.1%; Pred. No. 3.09e-06;	



W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 19:20:21 1999; MasPar time 112.95 Seconds
Tabular output not generated. 1187.596 Million cell updates/sec

Title: >US-08-702-718-3
Description: (1-1551) from US08702718.seq
Perfect Score: 1551
N.A. Sequence: 1 TCCTCTAACCTTGACCTTGG.....AAAAA.....TTTTTTTTTTTTTTTT
Comp: AGGAGATTGGAAGTGAAGC.....TTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 165359 seqs, 43243793 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCIT9_COMB 5:backfiles1
Statistics: Mean 8.940; Variance 5.499; scale 1.626

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description	Pred. No.
C 1	48	3.1	7218	2	US-08-232- Sequence 14, Applicati	8.27e-14
2	35	2.3	215	1	US-08-238- Sequence 5, Applicatio	1.86e-06
3	32	2.1	654	4	PCT-US95-0 Sequence 11, Applicati	7.54e-05
4	32	2.1	940	3	US-08-471- Sequence 1, Applicatio	7.54e-05
5	32	2.1	965	3	US-08-388- Sequence 22, Applicati	7.54e-05
6	31	2.0	208	2	US-08-686- Sequence 37, Applicati	2.52e-04
7	31	2.0	1046	1	US-08-361- Sequence 4, Applicatio	2.52e-04
8	31	2.0	1046	3	US-08-484- Sequence 2, Applicatio	2.52e-04
9	31	2.0	1137	3	US-08-706- Sequence 1, Applicatio	2.52e-04
10	31	2.0	1856	3	US-08-706- Sequence 1, Applicatio	2.52e-04
11	31	2.0	2589	4	PCT-US96-1 Sequence 1, Applicatio	2.52e-04
12	31	2.0	5173	2	US-08-242- Sequence 1, Applicatio	2.52e-04
13	30	1.9	41	1	US-08-113- Sequence 39, Applicati	8.31e-04
14	30	1.9	41	1	US-08-113- Sequence 41, Applicati	8.31e-04
15	30	1.9	55	1	US-08-113- Sequence 41, Applicati	8.31e-04
16	29	1.9	92	1	US-08-120- Sequence 94, Applicati	2.70e-03
17	29	1.9	92	1	US-08-478- Sequence 94, Applicati	2.70e-03
18	29	1.9	97	3	US-08-478- Sequence 87, Applicati	2.70e-03
19	29	1.9	97	1	US-08-120- Sequence 87, Applicati	2.70e-03
20	29	1.9	140	1	US-08-628- Sequence 5, Applicatio	2.70e-03

21	29	1.9	240	1	US-08-628- Sequence 6, Applicatio	2.70e-03
22	29	1.9	358	2	US-08-686- Sequence 3, Applicatio	2.70e-03
23	29	1.9	531	4	PCT-US95-0 Sequence 19, Applicati	2.70e-03
24	29	1.9	575	3	US-08-534- Sequence 11, Applicati	2.70e-03
25	30	1.9	684	3	US-08-226- Sequence 27, Applicati	8.31e-04
26	30	1.9	961	5	5219739-16 Patent No. 5219739.	8.31e-04
27	30	1.9	961	5	5194596-16 Patent No. 5194596.	8.31e-04
28	29	1.9	965	3	US-08-388- Sequence 22, Applicati	2.70e-03
29	29	1.9	1026	3	US-08-713- Sequence 6, Applicatio	2.70e-03
30	29	1.9	2223	3	US-08-724- Sequence 1, Applicatio	2.70e-03
31	29	1.9	2678	3	US-08-724- Sequence 2, Applicatio	2.70e-03
32	29	1.9	2989	5	5378464-1 Patent No. 5378464.	2.70e-03
33	29	1.9	5852	1	US-07-867- Sequence 2, Applicatio	2.70e-03
34	29	1.9	6877	1	US-08-347- Sequence 1, Applicatio	2.70e-03
35	28	1.8	672	2	US-07-847- Sequence 21, Applicati	8.66e-03
36	28	1.8	688	5	5498694-3 Patent No. 5498694.	8.66e-03
37	28	1.8	1243	1	US-08-178- Sequence 1, Applicatio	8.66e-03
38	28	1.8	1264	3	US-08-739- Sequence 13, Applicati	8.66e-03
39	28	1.8	1283	1	US-08-174- Sequence 19, Applicati	8.66e-03
40	28	1.8	1493	1	US-08-340- Sequence 24, Applicati	8.66e-03
41	28	1.8	1602	2	US-08-530- Sequence 3, Applicatio	8.66e-03
42	28	1.8	2760	1	US-08-101- Sequence 3, Applicatio	8.66e-03
43	28	1.8	3095	5	5231168-1 Patent No. 5231168.	8.66e-03
44	28	1.8	3637	2	US-08-445- Sequence 3, Applicatio	8.66e-03
45	28	1.8	3973	1	US-08-036- Sequence 21, Applicati	8.66e-03

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx
DT Sequence 14, Application US/08232463
CC Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935.313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: PT29pt-F1s
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 3.1%; Score 48; DB 2; Length 7218;
Best Local Similarity 0.6%; Pred. No. 8.27e-14;
Matches 2; Conservative 195; Mismatches 149; Indels 0; Gaps 0;

Db 1065 ATYY 1124
Cc 1428 ATGTTTCTGAGTCCTCGTATCTCTCATCTCGGAATAAATCTTTCTGTCCTCAAA 1369
Db 1125 YY 1184
Cc 1368 CTATATTATGTAGATTCGACAAAGAACACAGTGAGTTGATATGTCATCAATGTTATGC 1309
Db 1185 YY 1244
Cc 1308 TCTCTTTTACAAAACCTTTTCAAGGCATCTCCATAGTACACTCTTGCGCTCTCTAGCGG 1249
Db 1245 YY 1304
Cc 1248 CAGCCAGAGCTCGTCCATATAAGCTGTGAGCATATCCAGAGACTCTCTTGATACCCC 1180
Db 1305 YY 1364
Cc 1188 AAACAAAGCGTATAGTATCTGCTCTCTCAACACATAGTGGTTCAGCAAAAGCTCCACT 1129
Db 1365 YY 1410
Cc 1128 ATGAGCATCAACATAGGCCATGATGTTCTTTACCTTTCCAGCTCT 1083

RESULT 2
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC xxxxx

DE Sequence 5, Application US/08238163
CC Sequence 5, Application US/08238163
CC Patent No. 5569830
CC GENERAL INFORMATION:
CC APPLICANT: BENNETT, Alan
CC APPLICANT: LABAVITCH, John M.
CC APPLICANT: POWELL, Ann
CC APPLICANT: STOTZ, Henrik
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL
CC TITLE OF INVENTION: POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend Khourie and Crew
CC STREET: Steuart Street tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/238,163
CC FILING DATE: 03-MAY-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc_feature
CC LOCATION: 1..215
CC OTHER INFORMATION: /standard_name= "Deduced amino acid
CC OTHER INFORMATION: sequence of PGIP from bean."
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 2.3%; Score 35; DB 1; Length 215;
Best Local Similarity 15.9%; Pred. No. 1.88e-06;
Matches 25; Conservative 67; Mismatches 63; Indels 2; Gaps 2;

Db 8 SSVVSRATASCDNRKAKDGTTSWTTDCCNRTWGVCDTDTTYRVNNDSGHNYSSANYN 67
Cc 507 GCGAACAATAGGAGCTCGCTTGTGATGCTCAAGTCTCTGTTGTCAGCTTA-TG 565
Db 68 YGNNVGAAKTHYTHNVGADSKTIVTDSYNASGTSNNGDGNRSGADSYGSSKKTAM 127
Cc 566 TTTATCGGAGGATGTATGAAGATGACACAGTAATACCGCTGGATG-ACCTCCCTTGATTAT 624
Db 128 TSNRTGTANNVDSNRMGDASVGSCKNKKHAKNS 164
Cc 525 GGTGGAATTCGACACACATGTTGGGATTGATAGCC 661

RESULT 3
ID PCT-US95-06406A-11 STANDARD; DNA; UNC; 654 BP.
AC xxxxxx

DE Sequence 11, Application PC/TUS9506406A
CC Sequence 11, Application PC/TUS9506406A
CC GENERAL INFORMATION:
CC APPLICANT: Janet D. Robishaw, Charles Kunsch
CC TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
CC TITLE OF INVENTION: Subunits
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET:
CC CITY:
CC STATE:
CC COUNTRY:
CC ZIP:
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
CC COMPUTER: IBM 486
CC OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/06406A
CC FILING DATE: Herewith
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE:
CC TELEFAX:
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 654

CC GENERAL INFORMATION:
CC APPLICANT: Jacobs, Kenneth
CC APPLICANT: McCoy, John
CC APPLICANT: LaValle, Edward
CC APPLICANT: Racie, Lisa
CC APPLICANT: Merberg, David
CC APPLICANT: Treacy, Maurice
CC APPLICANT: Evans, Cheryl
CC APPLICANT: Spaulding, Vikki
CC TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
CC NUMBER OF SEQUENCES: 71
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc
CC STREET: 87 CambridgePark Drive
CC CITY: Cambridge
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02140
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/686,878A
CC FILING DATE:
CC CLASSIFICATION: 536
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brown, Scott A.
CC REGISTRATION NUMBER: 32,724
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 498-8224
CC TELEFAX: (617) 876-5851
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 208 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC SEQUENCE 208 BP: 109 A: 12 C: 23 G: 43 T: 21 OTHER.

Query Match 2.0%; Score 31; DB 2; Length 208;
Best Local Similarity 61.8%; Pred. No. 2.52e-04;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Db 87 GTAATTTGTCNGNGATAATTGAANATAACNGNGATTAAAGNNTNATGNTAAAAAAA 146
QY 1476 GTAATAATATTGTGTTCTTGATCAAAACATGAGATAAGAGTTTTCATAAAAAAAA 1535
Db 147 AAAAAAAAAAAAAA 162
QY 1536 AAAAAAAAAAAAAA 1551

RESULT 7
ID US-08-361-467B-4 STANDARD; DNA; UNC; 1046 BP.
AC xxxxxx
DT

DE Sequence 4, Application US/08361467B
CC Sequence 4, Application US/08361467B
CC Patent No. 5633441
CC GENERAL INFORMATION:
CC APPLICANT: De Greef, Willy
CC APPLICANT: Van Emmelo, John
CC APPLICANT: De Oliveria, Dulce E.
CC APPLICANT: De Souza, Maria-Helena
CC APPLICANT: Van Montagu, Marc
CC TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
CC STREET: P O Box 1404
CC CITY: Alexandria
CC STATE: Virginia
CC COUNTRY: United States
CC ZIP: 22313-1404
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/361,467B
CC FILING DATE: 22-DEC-1994
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/681,492
CC FILING DATE: 04-APR-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/EP90/01275
CC FILING DATE: 01-AUG-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 89 402 224.3
CC FILING DATE: 04-AUG-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Schulman, Robert M
CC REGISTRATION NUMBER: 31,196
CC REFERENCE/DOCKET NUMBER: 010830-027
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 836-6620
CC TELEFAX: (703) 836-2021
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1046 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA to mRNA
CC IMMEDIATE SOURCE:
CC CLONE: 3C9
CC SEQUENCE 1046 BP: 360 A: 254 C: 166 G: 266 T: 0 OTHER.

Query Match 2.0%; Score 31; DB 1; Length 1046;
Best Local Similarity 60.8%; Pred. No. 2.52e-04;
Matches 87; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Db 872 AAGCAAGAGGAAAGAGCTAGTTTCTTGTTCCTTATTTTCTCTCTATCAAAAC 931
QY 1409 AAGCAGGAGCTCAGAACATAGTTTCTTGTCTGCTGAGGTTTCGGTTTATAT 1468
Db 932 CCAACAAGTAAATGGATTATAAGTTTCTTCAAAAAAAAAAAAAAAAAAAAA 991
QY 1469 TTCCTTCTAAATATATTGTATGTTTCTTCTGATCAAAACATGAGATAAGAGTTTTCATA 1528
Db 992 AAAAAAAAAAAAAAAAAAAAAA 1014
QY 1529 AAAAAAAAAAAAAAAAAAAAAA 1551

RESULT 8
ID US-08-484-332C-4 STANDARD; DNA; UNC; 1046 BP.
AC xxxxxx
DT

DE Sequence 4, Application US/08484332C
CC Sequence 4, Application US/08484332C
CC Patent No. 5767374
CC GENERAL INFORMATION:
CC APPLICANT: De Greef, Willy
CC APPLICANT: Van Emmelo, John
CC APPLICANT: De Oliveria, Dulce E.
CC APPLICANT: De Souza, Maria-Helena
CC APPLICANT: Van Montagu, Marc
CC TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR

CC	TITLE OF INVENTION:	EMBRYOS
CC	NUMBER OF SEQUENCES:	13
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	BURNS, DOANE, SWECKER & MATHIS
CC	STREET:	P.O. Box 1404
CC	CITY:	Alexandria
CC	STATE:	Virginia
CC	COUNTRY:	United States
CC	ZIP:	22313-1404
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/484,332C
CC	FILING DATE:	7-JUNE-1995
CC	CLASSIFICATION:	800
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US 08/361,467
CC	FILING DATE:	22-DEC-1994
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US 07/681,492
CC	FILING DATE:	04-APR-1991
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	WO PCT/EP90/01275
CC	FILING DATE:	01-AUG-1990
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	EP 89 402 224.3
CC	FILING DATE:	04-AUG-1989
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Schulman, Robert M.
CC	REGISTRATION NUMBER:	31,196
CC	REFERENCE/DOCKET NUMBER:	010830-093
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(703) 836-6620
CC	TELEFAX:	(703) 836-2021
CC	INFORMATION FOR SEQ ID NO:	4:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	1046 base pairs
CC	TYPE:	nucleic acid
CC	STRANDEDNESS:	double
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	cDNA to mRNA
CC	IMMEDIATE SOURCE:	
CC	CLONE:	3C9
CC	SEQUENCE	1046 bp; 360 A; 254 C; 166 G; 266 T; 0 OTHERS
Db	Query Match	2.08; Score 31; DB 3: Length
	Best Local Similarity	60.88; Pred. No. 2,52e-04;
	Matches	87; Conservative 0; Mismatches 56; In
Db	872	AAGCAAGAGGAAAGAACGTAGTTTCTTCCTATTTCTCTATTTTGTTTC
Qy	1409	AAGCGAGACTCAGAACATAGTTTTCTTTGCTCTCTGCTGAGGTTT
Db	932	CCAAACAAGTAAGATGGATTATTAAGTTTTCCTCAAATAAAAAA
Qy	1469	TTCACTTGTAAATATATGTGATGGTTCTTGATCAAACATGAGATA
Db	992	AAAAAAAAAAAAAAAAAAAAA 1014
Qy	1529	AAAAAAAAAAAAAAAAAAAAA 1551
RESULT	9	
ID	US-08-706-214-2	STANDARD; DNA; UNC: 1137 BP.
AC	xxxxxx	
DT		
DE	Sequence 2,	Application US/08706214
CC	Sequence 2,	Application US/08706214
CC	Patent No. 5770720	
CC	GENERAL INFORMATION:	

```

CC APPLICANT: Deul, Thomas F.
CC APPLICANT: wang, Zhao-Yi
CC APPLICANT: Shenk, Thomas E.
CC TITLE OF INVENTION: UBIQUITIN CONJUGATING ENZYMES HAVING
CC TITLE OF INVENTION: UBIQUITIN CONJUGATING ENZYMES HAVING
CC TITLE OF INVENTION: TRANSCRIPTIONAL REPRESSOR ACTIVITY
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Paul A. Stone, Esq.
CC STREET: One Metropolitan Square, 16th Floor
CC CITY: St. Louis
CC STATE: Missouri
CC COUNTRY: USA
CC ZIP: 63102
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/706,214
CC FILING DATE: 30-AUG-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/018,040
CC FILING DATE: 21-MAY-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/002,995
CC FILING DATE: 30-AUG-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Stone, Paul A.
CC REGISTRATION NUMBER: 38,628
CC REFERENCE/DOCKET NUMBER: JWH 10033
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (314) 231-5400
CC TELEFAX: (314) 231-4342
CC TELEX: 6502697583 MCI
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1137 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: CDNA to mRNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 88..564
CC SEQUENCE 1137 BP; 294 A; 285 C; 275 G; 283 T; 0 OTHER.
SQ
Query Match 2.0%; Score 31; DB 3; Length 1137;
Best Local Similarity 89.7%; Pred. No. 2.52e-04;
Matches 35; Conservative 0; Mismatches 4; Indels 0
Db 1090 ATAAACTGTTTATATAAAAAAAAAAAAAAAAAAAAA 1128
QY 1513 ATAAAGAGTTTCTATAAAAAAAAAAAAAAAAAAAAA 1551
RESULT 10
ID US-08-706-214-1 STANDARD; DNA: UNC; 1856 BP.
AC *****
DT
DE Sequence 1, Application US/08706214
CC Sequence 1, Application US/08706214
CC Patent No. 5770720
CC GENERAL INFORMATION:
CC APPLICANT: Deul, Thomas F.
CC APPLICANT: Wang, Zhao-Yi
CC APPLICANT: Shenk, Thomas E.
CC TITLE OF INVENTION: UBIQUITIN CONJUGATING ENZYMES HAVING
CC TITLE OF INVENTION: TRANSCRIPTIONAL REPRESSOR ACTIVITY
CC NUMBER OF SEQUENCES: 4

```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Paul A. Stone, Esq.
CC STREET: One Metropolitan Square, 16th Floor
CC CITY: St. Louis
CC STATE: Missouri
CC COUNTRY: USA
CC ZIP: 63102
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/706,214
CC FILING DATE: 30-AUG-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/018,040
CC FILING DATE: 21-MAY-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/002,995
CC FILING DATE: 30-AUG-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Stone, Paul A.
CC REGISTRATION NUMBER: 38,628
CC REFERENCE/DOCKET NUMBER: JWH 10033
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (314) 231-5400
CC TELEFAX: (314) 231-4342
CC TELEX: 6502697583 MCI
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1856 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: CDNA to mRNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 807..1283
CC SEQUENCE 1856 BP; 401 A; 499 C; 518 G; 438 T; 0 OTHER.

Query Match 2.0%; Score 31; DB 3; Length 1856;
Best Local Similarity 89.7%; Pred. No. 2.52e-04;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1809 ATAACTGTTTATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1847
Qy 1513 ATAAAGAGTTTTCATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1551

RESULT 11
ID PCT-US96-12860-1 STANDARD; DNA; UNC: 2589 BP.
AC xxxxxx
DT
DE Sequence 1, Application PC/TUS9612860
CC Sequence 1, Application PC/TUS9612860
CC GENERAL INFORMATION:
CC APPLICANT: TULARIK, INC.
CC TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOHRBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/12860
CC FILING DATE: 06 AUG 1996
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brenner, David J.
CC REGISTRATION NUMBER: 24,774
CC REFERENCE/DOCKET NUMBER: A-62464/DJB
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415)781-1989
CC TELEFAX: (415)398-3249
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2589 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: CDNA
CC SEQUENCE 2589 BP; 844 A; 436 C; 522 G; 787 T; 0 OTHER.

Query Match 2.0%; Score 31; DB 4; Length 2589;
Best Local Similarity 89.7%; Pred. No. 2.52e-04;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 2520 ATAAAGTCTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2558
Qy 1513 ATAAAGAGTTTTCATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1551

RESULT 12
ID US-08-242-677-1 STANDARD; DNA; UNC: 5173 BP.
AC xxxxxx
DT
DE Sequence 1, Application US/08242677
CC Sequence 1, Application US/08242677
CC Patent No. 5677143
CC GENERAL INFORMATION:
CC APPLICANT: Gaynor, Richard B
CC TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
CC TITLE OF INVENTION: and Uses Thereof in regulating Gene Expression and in th
CC TITLE OF INVENTION: Treatment of AIDS
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Arnold, White & Durkee
CC STREET: P.O. Box 4433
CC CITY: Houston
CC STATE: TX
CC COUNTRY: USA
CC ZIP: 77210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/242,677
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mayfield, Denise L.
CC REGISTRATION NUMBER: 33,732
CC REFERENCE/DOCKET NUMBER: UTSD:401
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 713-787-1400
CC TELEFAX: 713-789-2679
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:

```

CC LENGTH: 5173 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC FEATURE:
CC NAME/KEY: CDS
CC LOCATION: 1..4863
CC SEQUENCE 5173 BP; 1348 A; 1116 C; 1309 G; 1400 T; 0 OTHER.

Query Match      2.0%; Score 31; DB 2; Length 5173;
Best Local Similarity 69.1%; Pred. No. 2.52e-04;
Matches 56; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Db 5047 CACTTTTGTAAATTTGTTCTTCTTTAATAAATAATTTTAAGCAATTTGCCAATAAA 5105
      ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Qy 1471 CACTTCTAAATATATTGTATGGTTCTTGATCAAAACATGAGATAAAGAGTTTTCATAAA 1530
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 5107 AAAAAAAAAAAAAAAAAAAAAA 5127
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1531 AAAAAAAAAAAAAAAAAAAAAA 1551

RESULT 13
ID US-08-113-646A-39 STANDARD; DNA; UNC; 41 BP.
AC xxxxxx
DT
DE Sequence 39, Application US/08113646A
CC Sequence 39, Application US/08113646A
CC Patent No. 5578468
CC GENERAL INFORMATION:
CC APPLICANT: PICKUP, David J.
CC APPLICANT: PATEL, Dhaval Kumar
CC APPLICANT: ANTCHAK, James B.
CC TITLE OF INVENTION: SITE-SPECIFIC RNA CLEAVAGE
CC NUMBER OF SEQUENCES: 44
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 NORTH GLEBE ROAD
CC CITY: ARLINGTON
CC STATE: VIRGINIA
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/113,646A
CC FILING DATE: 31-AUG-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 1579-20
CC FILING DATE: 10-AUG-1987
CC ATTORNEY/AGENT INFORMATION:
CC NAME: WILSON, MARY J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 1579-20
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 39:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 41 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: RNA (genomic)
CC SEQUENCE 41 BP; 33 A; 2 C; 1 G; 0 T; 5 OTHER.

Query Match      1.9%; Score 30; DB 1; Length 41;
Best Local Similarity 69.1%; Pred. No. 8.31e-04;
Matches 29; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 4 AAAAAAAAAAAAAAAAAAAAAA 39
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1516 AAGAGTTTTCATATAAAAAAAAAAAAAAAAAAAAAA 1551
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
ID US-08-113-646A-41 STANDARD; DNA; UNC; 55 BP.
AC xxxxxx
DT
DE Sequence 41, Application US/08113646A
CC Sequence 41, Application US/08113646A
CC Patent No. 5578468

```

W P S R E H
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Sat Oct 23 18:29:26 1999; MasPar time 2634.07 Seconds
 1379.705 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-702-718-3
 Description: (1-1551) from US08702718.seq
 Perfect Score: 1551
 N.A. Sequence: 1 TCTCTTAACCTTGACCTTCG.....AAAAA.....AAAAA 1551
 Comp: AGGAGATTGGAATGCAAGC.....TTTTTTTTTTTTTTTT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0
 Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-est58
 1:em-est10 2:em-est11 3:em-est17 4:em-est18 5:em-est2
 6:em-est9 7:em-gss1
 Database: genbank-est11
 8:gb-est1 9:gb-est10 10:gb-est11 11:gb-est12 12:gb-est13
 13:gb-est14 14:gb-est15 15:gb-est16 16:gb-est17
 17:gb-est18 18:gb-est19 19:gb-est20 20:gb-est21
 21:gb-est22 22:gb-est23 23:gb-est24 24:gb-est25
 25:gb-est26 26:gb-est27 27:gb-est28 28:gb-est29
 29:gb-est30 30:gb-est31 31:gb-est32 32:gb-est33 33:gb-est34
 34:gb-est35 35:gb-est36 36:gb-est37 37:gb-est38 38:gb-est39
 39:gb-gss3 40:gb-gss4 41:gb-gss5 42:gb-gss6

Statistics: Mean 12.169; Variance 6.306; scale 1.930

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	281	18.1	471	20	AA896953 L30-386T3 Ice plant La	1.08e-189
2	223	14.4	502	17	T44185 7448 Lambda-PRL2 Arabi	2.47e-143
3	221	14.2	478	31	R90544 16899 Lambda-PRL2 Arab	9.57e-142
4	201	13.0	676	23	EST220465 Normalized r	5.60e-126
5	191	12.3	657	15	AA606966 vm944009.r1 Knowles Sol	5.06e-118
6	188	12.1	417	24	C96653 Rice callus Ory	1.16e-115
7	183	11.8	527	31	R90561 16916 Lambda-PRL2 Arab	9.83e-112
8	178	11.5	620	26	AU002699 Bombyx mori p	8.20e-108
9	164	10.6	501	11	AA313713 EST185581 Colon carcin	7.09e-97
10	156	10.1	522	10	AA237580 mx11d01.r1 Soares mous	1.17e-90

11	151	9.7	307	10	AA231705	CD0534.R cDNA from oat	8.76e-87
12	143	9.2	466	36	AA112941	zn59b06.r1 Stratagene	1.31e-80
13	143	9.2	473	14	AA498291	vh38q11.r1 Barstead mo	1.31e-80
14	135	8.7	213	8	Z17455	ATIS0029 AC16H Arabido	1.83e-74
15	135	8.7	252	17	AA720224	33417 Lambda-PRL2 Arab	1.83e-74
16	125	8.1	338	11	AA337465	EST42223 Endometrial t	7.88e-67
17	123	7.9	494	18	AA830955	oc58d01.s1 NCI-CGAP GC	2.61e-65
18	121	7.8	489	23	AJ168709	oc56a01.s1 Soares_NHMH	8.57e-64
19	119	7.7	321	19	F11594	HSC22A101 normalized i	2.80e-62
20	119	7.7	418	34	W40867	mc56e02.r1 Soares mous	2.80e-62
21	120	7.7	442	33	W01297	yz94h09.r1 Soares mela	4.90e-62
22	119	7.7	444	10	AA245100	mx08g10.r1 Soares mous	2.80e-62
23	119	7.7	503	9	AA178724	mt16h01.r1 Soares mous	2.80e-62
24	119	7.7	509	9	AA177425	mt23d11.r1 Soares mous	2.80e-62
25	116	7.5	337	11	AA321974	EST24544 Cerebellum II	5.17e-60
26	117	7.5	343	21	AA992542	ot88g03.s1 Soares-tota	9.09e-61
27	117	7.5	365	11	AA25019	EST27941 Cerebellum II	9.09e-61
28	117	7.5	452	9	AA169633	zq59h08.r1 Stratagene	9.09e-61
29	117	7.5	510	21	C88687	C88687 Mouse early bla	9.09e-61
30	114	7.4	331	11	AA332541	EST36505 Embryo, 8 wee	2.93e-59
31	114	7.4	506	26	AA901227	UI-R-A1-dp-g-06-0-UI.s	1.66e-58
32	114	7.4	540	24	A1338915	GH15054.5prine GH Dros	1.66e-58
33	114	7.4	752	25	A1327328	mp70f10.x1 Soares 2NDM	1.66e-58
34	112	7.2	191	34	W49861	zEST00324 Maize Leaf,	5.31e-57
35	109	7.0	455	13	AA427173	ve83c01.r1 Soares mous	9.46e-55
36	109	7.0	461	9	AA196934	zq59h09.r1 Stratagene	9.46e-55
37	107	6.9	425	33	W14146	mb19e03.r1 Soares mous	2.97e-53
38	105	6.8	302	12	AA356125	EST64627 Jurkat T-cell	9.23e-52
39	106	6.8	358	32	N36726	YX90C03.r1 Soares mela	1.66e-52
40	104	6.7	360	9	AA205271	zq71c12.r1 Stratagene	5.13e-51
41	102	6.6	585	22	A1054878	coau0002G14 Cotton Bol	1.58e-49
42	101	6.5	331	16	AA682985	ae89c01.s1 Stratagene	8.72e-49
43	101	6.5	566	15	AA568724	nm21f09.s1 NCI-CGAP Co	8.72e-49
44	100	6.4	299	11	AA300831	EST13692 Testis tumor	4.81e-48
45	98	6.3	524	25	A1324744	mb82a02.x1 Soares mous	1.45e-46

ALIGNMENTS

RESULT 1
 LOCUS AA896953 471 bp mpna EST 06-APR-1998
 DEFINITION L30-386T3 Ice plant Lambda Uni-zap XR expression library, 30 hours
 NACI treatment Mesembryanthemum crystallinum cDNA clone L30-386 5'
 similar to Citrate synthase, mitochondrial precursor, mRNA
 sequence.
 ACCESSION AA896953
 NID g3033346
 VERSION AA896953.1 GI:3033346
 KEYWORDS EST.
 SOURCE common ice plant.
 ORGANISM Mesembryanthemum crystallinum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
 Cushman, J.C.
 REFERENCE 1 (bases 1 to 471)
 AUTHORS An expressed sequence tag database for the common ice plant,
 TITLE Mesembryanthemum crystallinum
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150267.

Contact: Cushman JC
 Department of Biochemistry and Molecular Biology
 Oklahoma State University
 350 Noble Research Center, Stillwater, OK 74078-3035, USA
 Tel: 405-744-5207
 Fax: 405-744-7799
 Email: jcushman@biochem.okstate.edu
 PCR Primers
 FORWARD: T7
 BACKWARD: T3
 Plate: L30-4 row: H column: 6
 Seq primer: T3

High quality sequence stop: 320.

On Jan 7, 1998 this sequence version replaced gi:948518.

```

FEATURES
  source
    Location/Qualifiers
      1..471
        /organism="Mesembryanthemum crystallinum"
        /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI"
        /db_xref="taxon:3544"
        /clone="L30-386"
        /clone_lib="Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment"
        /tissue_type="Leaf, 30 h 0.4M NaCl"
        /dev_stage="Six week old"
BASE COUNT 127 a 95 c 122 g 127 t
ORIGIN
  Query Match 18.1%; Score 281; DB 20; Length 471;
  Best Local Similarity 84.3%; Pred. No. 1.08e-189;
  Matches 376; Conservative 0; Mismatches 65; Indels 5; Gaps 5;
  Db 1 GCCATTGGTGGTACGACATTCAGATCCCTTACCTTTCATTTCGACGTCGACTGAATG 60
  QY 740 GCCATTGGTGGTACGACATTCAGATCCCTTACCTTTCATTTCGACGTCGACTGAATG 799
  Db 61 GGTGGTGGCCCACTTCATGATTAGCCAAATCAGGAAGTGTGCTGSGATCAAGTCAG 120
  QY 800 GTTGGTGGCCCACTTCATGATTAGCCAAATCAGGAAGTGTGCTGSGATTAATCAAG 859
  Db 121 TTGTGATGAATGTGAGAAACATCAGCAGGACGACCTAAAGATTATGCTGGAAGA 180
  QY 860 TTGTGATGAATGTGAGAGACATCTCGACAGACGATTTGAAGATTATGTTGGAAGA 919
  Db 181 CATTAAAGAGTGAAGGTTGTACCTGGATATGGCCATGGAGTTCTGGCAGACGGATC 240
  QY 920 CACTAAACAGTGCAGAGTTGTACCTGGATATGGCCATGGAGTTCTGGCAGACGGATC 979
  Db 241 CAGATATCTTCCGACAGAGGACTTTGCCCTTAAGCACTTACCTAATGATCCACTATTTC 300
  QY 980 CAAGATACATATGCAAGAGAGATTGTGTTGAAGCACTTTCCTGATGACCACTTTTTC 1039
  Db 301 AACTGTTTTCGAAGCTGTATGAA-TGGTGCTCCCAATTCGACAGAGCTTGGGAAGTAA 359
  QY 1040 AATTGTTGCAAGTTGTATGAAGTGGTGCTCCCTTCTATTAGAGCTTGGGAAGTAA 1099
  Db 360 AGAA-CAATGGCCGAATGTGATGCCACAGTGGGTGATTGCTGAACACTACTATGTTG 418
  QY 1100 AGAATCCATGGCCTAATGTGATGCTCATAGTGA-GTTTTGCTCAACCACTATGTTG 1158
  Db 419 ACAAGAAACAAGATACTACACTGTT 444
  QY 1159 ACA-GAA-GCAAGATACTATACGGTT 1182
  RESULT 2
  LOCUS 744185 502 bp mRNA EST 07-JAN-1998
  DEFINITION 7448 Lambda-PRL2 Arabidopsis thaliana cDNA clone 125C917, mRNA
  sequence.
  ACCESSION T44185
  NID 92758988
  VERSION 144185.1 GI:2758988
  KEYWORDS EST.
  SOURCE thale cress.
  ORGANISM Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
    Rosidae; Capparales; Brassicaceae; Arabidopsis.
  REFERENCE 1 (bases 1 to 502)
    Newman.T., deBruijn.F.J., Green.P., Keegstra.K., Kende.H.,
    McIntosh.L., Ohlrogge.J., Raikhel,N., Somerville.S., Thomashow.M.,
    Retzel.E. and Somerville.C.
    Genes galore: a summary of methods for accessing results from
    large-scale partial sequencing of anonymous Arabidopsis cDNA clones
    Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE 95148729
  RESULT 3
  LOCUS R90544 478 bp mRNA EST 30-DEC-1997
  DEFINITION 16899 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189D15T7, mRNA
  sequence.
  ACCESSION R90544

```

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9188
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: T7.

FEATURES
Location/Qualifiers
1..502
/organism="Arabidopsis thaliana"
/strain="var columbia"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
/db_xref="taxon:3702"
/clone="L23C97"
/clone_lib="Lambda-PRL2"

BASE COUNT 125 a 90 c 118 g 149 t 20 others
ORIGIN

Query Match 14.4%; Score 223; DB 17; Length 502;

Best Local Similarity 78.2%; Pred. No. 2.47e-143;

Matches 333; Conservative 0; Mismatches 88; Indels 5; Gaps 5;

Db 10 GATTATGTTGCAAAATTTTCCCAATGTTGGGATTTGATGATGAAAGGTCGAAAGAGCTC 69

QY 619 GATTATGTTGCAAAATTTTCCCAATGTTGGGATTTGATGATGAAAGGTCGAAAGAGCTC 678

Db 70 ATGAGGCTTTACATCACCATCCACAGTATCATGAAGTGAATGTTAGTGCTCACACT 129

QY 679 ATGCGCTTTATGTCACAAATTCACAGTATCATGAAGTGAATGTTAGTGCTCACACT 738

Db 130 GGTACCTGGTGTGTTAGTGCACCTTTACAGACCCATATCTGTCATTTCGAGCTGCAATTAAT 189

QY 739 GCCATTTGGTGGGTAGTCCACTTTTCAGATCCTTATTGTCATTTCGAGCAGCATTAAT 798

Db 190 GGTTAGCTGGCCCACTCCATGGTTTGGCTAATCAGGAAGTTTTCGTTTGGATCAATCA 249

QY 799 GGTGGCTGGCCCACTCCATGGATTAGCCAAACAGGAGTCCCTGCTGTGATTAATCA 858

Db 250 GTCGTAGAGGAATGTGGAGAAATATATCAAAAGAACAGTTGAAAGAAATATGTTGGAA 309

QY 859 GTTGTGATGAATGTGGAGAAACATCTCGACAGACAGTTCGAAGATTATGTTGGAAG 918

Db 310 ACATTAACAGTGGCAAGGTTATTCGCGGGTATGTCACGGTGTTCNCGCAATACTTGA 369

QY 919 ACATTAACAGTGGCAAGGTTATTCGCGGGTATGTCACGGTGTTCNCGCAAGACA-GA 977

Db 370 TCCCGCAATTTGTTGCCAAAGAGAAATTTCCNTNGNAGNANCTTCCCTNANGGCCCC 429

QY 978 TCCAAGATACAT-GCCAAAGAG-AAITTCG-GTTGAAGCA-CTTGCTGATGACCCAT 1033

Db 430 TTTTTC 435

QY 1034 TTTTTC 1039

NID 9958084
 VERSION R90544.1 GI:958084
 KEYWORDS EST.
 SOURCE
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Rosidae; Capparales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 478)
 AUTHORS Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H.,
 McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
 Retzel,E. and Somerville,C.
 TITLE Genes galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 JOURNAL Plant Physiol. 106, 1241-1255 (1994)
 MEDLINE 95148729
 COMMENT On Apr 14, 1993 this sequence version replaced gi:693655.

Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313tcn@bm.cl.msu.edu
 Seq primer: 17 dye primer.

FEATURES
 source
 1..478
 /organism="Arabidopsis thaliana"
 /strain="var columbia"
 /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
 Lambda PRL2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dt primed cDNA."
 /db_xref="taxon:3702"
 /clone="189p157"
 /clone_lib="Lambda-PRL2"
 BASE COUNT 119 a 90 c 113 g 137 t 19 others
 ORIGIN

Query Match 14.2%; Score 221; DB 31; Length 478;
 Best Local Similarity 81.6%; Pred. No. 9,57e-142;
 Matches 302; Conservative 0; Mismatches 65; Indels 3; Gaps 3;
 Db 1 CTTTCAGACCCATATCTGTCATTTCGAGCTGCATTAAATGTTAGCTGGGCCACTCCAT 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 760 CTTTCAGATCCTATTGTTGTCATTTCGAGCAGCATTAAATGTTGCTGGGTCCTCCAT 819
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 61 GGTTCGGCTAATCAGAAAGTTTGCTTCGATCAATAGTCGTAGAGAGATGTGGAGAA 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 820 GGATTAGCCCAACAGAAATCTGCTGTCGATTAAATCATCAGTTGTTGATGAATGTGGAG 879
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 121 GATATATCAAAAGACAGTTGAAAGAAATATGTTTGAAGAAATTAACAGTGGCAAGTT 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 880 AACATCTGCACAGAGAGTTGAAGATTATGTTGGAAGACACTAAACAGTGGCAAGTT 939
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 181 ATTCGGGATATGTCACGGTGTCTGCGCAATCTGATCCAGATATGATGCCAAGA 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 940 GTACCTGGATTGGTCTAGGAGTATTGCGGAAGACAGATCCAGATACACATGCCAAGA 999
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 241 GAATTTGCTTGAAGCATCTACTGACGCCCTCTTTCCAGCTGGTGTCAAGCTTTAT 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 1000 GAATTTGCTTGAAGCATTGCGCTGATGACCCATTCTTTCAATTTGGTGTCAAGTTGAT 1059
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 301 GAAGTTGTGCTCCCTGTTCTCACTGAGCTTTGGAAGGTTGAAGAACCCGTTGGCCAAAT 360
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1060 GAAGTGTGCTCC-TATTCTATTAGACTT-GGAAA-GGTAAGAAATCATGCCTAAT 1116
 Db 361 NTTGATGCTC 370
 ||||| |||||
 QY 1117 GTTGATGCTC 1126
 ||||| |||||
 RESULT 4
 LOCUS A1176862 676 bp mRNA EST 20-JAN-1999
 DEFINITION EST220455 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
 ROVBX80 3' end, mRNA sequence.
 ACCESSION A1176862
 NID g3727500
 VERSION A1176862.1 GI:3727500
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 676)
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
 Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2152042.

Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21
 FEATURES
 source
 1..676
 /organism="Rattus sp."
 /note="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 /db_xref="ATCC (inhost):2031570"
 /db_xref="taxon:10118"
 /clone="ROVBX80"
 /clone_lib="Normalized rat ovary, Bento Soares"
 BASE COUNT 156 a 188 c 150 g 181 t
 ORIGIN

Query Match 13.0%; Score 201; DB 23; Length 676;
 Best Local Similarity 68.1%; Pred. No. 6.60e-126;
 Matches 377; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
 Db 123 CTTGGGCTTTCTAGAGGGAAGGCTAGGGCTCTGCTCCAGATAAGCTGGGCCAACACACC 182
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 1266 CTTTGGGCTCTTAGCGGCAAGCCAGAGCTCGTCCCATATAGCTGTGAGCATATCC 1207
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 183 CAGTGCCTCGACACTCCGAAACAGGACTGTGTAGTAATTCATCTCGTCAATGCAATGA 242
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 1206 AAGACTCCTTGATACCCCAACAAACCGATAGTATCTTCTGTCACCAAGCATAGTG 1147
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 243 CTGGAGCAGACCCCACTGTGAGCATCTAGCTGGGCCAGGGTCTCTTAGCCCTTCCTTG 302
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 1146 GTTCAGCAAACTCCATATGAGCATCAACATTAGGCCATGGATTCTTTACCTTTCCAAG 1087
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 303 CTCTAAGAGGATTTGGGCACAATCTTTGTACAGCTAGCAACCACTGTTAAACATGGGATC 362
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 1086 CTCTAATAGAATAGGAGGCCACCTTTCATCAACTTTGACACCAATTTGAAAAAATGGGTC 1027
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 363 CTTGGGAGATGTTTCAGACAAACTCTCCGTACACAGGAATATCGTGGATCAGTCTTCCT 422
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 1026 ATCAGCAAGTGTCTTAACGCAAAATCTCTTTGGCATGTGTATCTTTGGATCTGTCTTCG 967
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 423 TAGCACTGCATGACCATATCTGTGGACAACCGCTCCGAGTTGAGTGTGTTCAGATGTA 482
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 966 CAATACTCTAGACCAAAATCCAGGTACAACTTGCCACTGTTTAGTGCTTCCTCAAAACATA 907

Db 483 GTCTCGTAAGTCTCTATCTGACACAGCTCTTTGCCGACTTCCTCTGTAGCTCGGTACGCCA 542

Cp 906 ATCTTTCAACTGCTGTGCGAGATGTTCTCTCCACATTCATCAACAAGTGTATTAATCCA 847

Db 543 GACAACACCTCTCTGTTTCTAGTCTCATGAGAGGCGCTGCCAGCCATCATCATAGTGTGC 602

Cp 846 CACGAGACTTCTGTGTGGCTTAATCCATGGAGTGCGCCACCAACCATTAATGTGTC 787

Db 603 TGCANAGGACAGGTAAAGGTCTGAAGGGCGCTGCCACCAAGTGGCTTGTGTGGGCACT 662

Cp 786 TGCAATGACAAATAGGATCTGAAAGTGGACTACCCACCAATGGCCAGTGTGTGCACT 727

Db 663 TACATTGCCACCT 676

Cp 726 AACATTTCCACCT 713

RESULT 5

LOCUS AA606966 657 bp mRNA EST 30-SEP-1997

DEFINITION vm94h09_r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA clone IMAGE:1005953 5' similar to SW:CSI_PIG P00889 CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR ;, mRNA sequence.

ACCESSION AA606966

NID q2455859

VERSION AA606966.1 GI:2455859

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS 1 (bases 1 to 657)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellingberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1328460.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:570169

High quality sequence stop: 473.

Location/Qualifiers

1..657

/organism="Mus musculus"

/strain="B6D F1/J"

/note="Organ: embryo; Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGTCGACCGTCGACCGCTTTTCTTTT-3', cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."

/db_xref="taxon:10090"

/clone="IMAGE:1005953"

/clone.lib="Knowles Solter mouse blastocyst B1"

/tissue_type="blastocyst"

/dev_stage="embryo (pre-implantation)"

/lab_host="DH10B"

BASE COUNT 161 a 167 c 181 g 148 t

ORIGIN

Query Match 12.3%; Score 191; DB 15; Length 657;

Best Local Similarity 67.3%; Pred. No. 5.06e-118;

Matches 410; Conservative 0; Mismatches 195; Indels 4; Gaps 4;

Db 6 CCTCAGTTCACCGAGCTCATCGCTTTGTACCTCACCATCCATAGTAGACCATGAGGTGG 65

Qy 660 CCTCAGATGCTTGGAGCTGATGCGCCCTTTATGTCACAATTCACAGTATCATAGGAGT 714

Db 66 TAATTAAGTCCGCCACACAAAGCCATTTGGTGGGAGGCGCA-TTTCAGACCCCTTACCTGTC 124

Qy 720 AAATGTTTGTACACACATGCGCCATTTGGTGGGTAGTCCACATTTTCAGATCTTTATTGTC 779

Db 125 CTTTCACAGCATGAATGGCTGGCGGGCT-CTACATGGACTAGCAATCAGGAGGT 183

Qy 780 ATTTGCACGACATTAATGTTTGGCTGGCCACTCCATGGATTAGCCAAACAGGAGT 839

Db 184 GCTTGTCTGTGTGACACAGCTACAGAAAGAGTTGGCAAGAGTGTGCAAGAGTGTGCAAGAGT 243

Qy 840 CCTGCTGTGGATTAAATCAGTTGTTGATGAATGTGGAGAGAACATCTCGACAGAGCAGT 899

Db 244 ACGAGACTACATCTGGAAACACACTCAATTCAGAGCGGTGTCGCCAGATACGTCATGC 303

Qy 900 GAAACATTATGTTGGAGACACATAACAGTGGCAAGGTTGTACCTGGATTTGGTCTAGG 959

Db 304 AGTACTGAGGAAGACTGACCCCTCGCTATTCTGTGACGAGAGTTTGTCTGAAACATCT 363

Qy 960 AGTATTGCGGAAGACAGATCCAGATACACATGCAAAAGAGAAATTTGCGTTGAAGCATT 1019

Db 364 GCCTAAGGATTCCCATGTTCAAGCTGTGCTCAGCTGTACAGATGTGCGCAATATC 423

Qy 1020 GCGCTATG-ACCCATTTTTCAT-TGCTGTCAAAGTTGTATGAAGTGGTGCCTCTATT 1077

Db 424 CTCTTAGAGCAAGGAAGGCTTAAGAACCTTTGGCCCAACCTAGACGCTACACATGCGGTG 483

Qy 1078 CTATTAGAGCTTGGAAAGTAAACATCCATGGCTTAATGTTGATGCTCATAGTGAGTT 1137

Db 484 CTGCTCCAGTACTATGCGATACCGAGATGAACACTACTACAGTCCTGTTTGGAGTGTCT 543

Qy 1138 TTGCTGAACCACTATGTTTGCAGAGAACAGATACTATACGGTTTGTGTTGGGATCA 1197

Db 544 CGGGCAGCTGGGTGCTAGCCAGCTCATCTGGAGCAGAGCCCTAGGCTTCCCCCTGGAA 603

Qy 1198 AGGAGCTTGGAAATATGCTCAGAGCTTATGGACCGAGCTCTTGCTGCGCTAGAG 1257

Db 604 AGGCCCAAG 612

Qy 1258 AGGCCCAAG 1266

RESULT 6

LOCUS C96653 417 bp mRNA EST 19-OCT-1998

DEFINITION C96653 Rice callus Oryza sativa cDNA clone C10531_5A, mRNA sequence.

ACCESSION C96653

NID q3758954

VERSION C96653.1 GI:3758954

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

REFERENCE 1 (bases 1 to 417)

AUTHORS Sasaki, T. and Minobe, Y.

TITLE Rice cDNA from callus

JOURNAL Unpublished (1994)

COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = RGP

Location/Qualifiers

1. .417

/organism="Oryza sativa"

/strain="cultivar Nipponbare, sub_species Japonica"

/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:

NotI; cDNA prepared from rice callus mRNAs by using

oligo(dT) as a primer and ligating to the SalI-NotI site

of pBluescript II SK+ phagemid."

/db_xref="taxon:4530"

/map="11"

/clone="C10531_5A"

/clone.lib="Rice callus"

BASE COUNT 115 a 76 c 108 g 118 t

ORIGIN

Query Match 12.1%; Score 188; DB 24; Length 417;
 Best Local Similarity 73.1%; Pred. No. 1.16e-115;
 Matches 305; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

Db 2 TATGTTTACCGGAGATCTTCAAGATGGGAAACTATAGCAGCTGATATGCACTGGAC 61

Qy 562 TATGTTTATCGAGGATGTATAAGATGGACAGTAATACCGGTGGATGACTCCCTTGAT 621

Db 62 TATCAGCAAACTTTTCACACATCTTGGGTTCATGATCCCAAAATGCTCGAGTTGATG 121

Qy 622 TATGTTGAAATTTCCGACACATCTTGGGATTTGATAGCCCTCAGATGCTTGAGCTGATG 681

Db 122 CGACTATATGTGACATCCACACTGATCATGAGGTGGAAATGTCAGTCTCATCTGGA 181

Qy 682 CGCCTTATGTCACATTCACAGTATCATGAGGTGGAATGTATGTCACACTGGC 741

Db 182 CATCTGGTTGAAGTCTCTGTCAGATCCCTTATCTTTTTCAGCTGCACTGAATGGT 241

Qy 742 CATTTGGTGGTATCCACTTTTCAGATCCCTTATTTGTCATTTGCAGCAGCATTAATGGT 801

Db 242 TTAGCTGACCATTCGAGGCTGGCTGCTAATCAGGAAGTGTGTGGATCAATCTGTA 301

Qy 802 TTGCTGGCCACTCCATGGATTAGCAACAGGAAGTCTGCTGTGGATTAATCAGTT 861

Db 302 ATAGTGTGACAGACAGTAGTATGTTTACACTGATCAACTCAAAGAGATGTGTGGAAGACA 361

Qy 862 GTTGATGAATCTGGAGAGAACATCTGCACAGAGCAGTTGAAAGATTATGTTGGAAGACA 921

Db 362 CTAAAAGTGAAGAGTTGTCTCTGGCTC-GGTCATGGAGTTCTACGTAAGACCGAT 417

Qy 922 CTAAACAGTGAAGGTTGTACCTGGATTTGGTCTAGGAGTATTGCGGAAGACAGAT 978

RESULT 7 R90561 527 bp mRNA EST 30-DEC-1997
 LOCUS 18916 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189H3T7, mRNA
 DEFINITION sequence.

ACCESSION R90561

NID 958101

VERSION R90561.1 GI:958101

KEYWORDS EST

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 527)

AUTHORS Newman, I., deRuiz, F.J., Green, P., Keegstra, K., Kende, H.,

McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,

Retzel, E., and Somerville, C.

TITLE Large-scale partial sequencing of anonymous Arabidopsis cDNA clones

JOURNAL Plant Physiol. 106, 1241-1255 (1994)

95148729

MEDLINE

On Apr 14, 1993 this sequence version replaced gi:693674.

COMMENT

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tcn@bm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1. .527

/organism="Arabidopsis thaliana"

/strain="var columbia"

/note="Vector: lambda zip-Lox; Site_1: SalI; Site_2: NotI;

Lambda PRL2 is a cDNA library derived from equal

quantities of 4 pools of mRNA. The mRNA sources were 1) 7

day germinated etiolated seedlings; 2) tissue culture

grown roots; 3) staged plants half with 24 hour light

cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and

silicles. The vector is BRU's lambda zip-Lox. The cDNA

inserts were directionally cloned with Sal-Not arms using

oligo dT primed cDNA."

/db_xref="taxon:3702"

/map="873F09; lq21.3-lq23.2; 9"

/clone="189H3T7"

/clone.lib="Lambda-PRL2"

BASE COUNT 130 a 106 c 122 g 148 t 21 others

ORIGIN

Query Match 11.8%; Score 183; DB 31; Length 527;
 Best Local Similarity 77.6%; Pred. No. 9.83e-112;
 Matches 298; Conservative 0; Mismatches 80; Indels 6; Gaps 5;

Db 3 ATTCTGTAGTCGACAGAAATATTACTACTGCCAGTCTGGAGTGAACCATTCACGGAG 62

Qy 211 ATACCTGAATGCCAGAACTTTTACCCGCTGCAAGTGTGGTCAGAGCCATTGCTGAA 270

Db 63 GGTCTTTGTGCTCTCTTTAACTGGAAGTGTGCTAGCAAGAGCAAGTTGAAGCACTG 122

Qy 271 GGTCTCTTTGCTCTCTTTTAAACCGGAAAGTTCTTAGCAAGAGCAAGTAGTCTCTA 330

Db 123 TCGAAGACTTGGCAACGCTGCTGTGCGAGATTATGCTACAATGCCATCGATGCT 182

Qy 331 TCAGCAGATTTACGAAAACGCTCTCTATCCAGACCATGTTGACAAACAATTTGCT 390

Db 183 CTGCTTCCACAGCTCATCCCAATGACTCAATTTGCTAGCGGTGTTATGGCCCTCCAGGTG 242

Qy 391 CTACCTATTACGGCTCATCCCAATGACTCAGTTTTCGACTGGTTATGCGCTTACAGACT 450

Db 243 CAAAGTGAGTTTCAAAGGCATATGAGATGGAATTCATAGTCAAAGTTCTGGAGGCA 302

Qy 451 CGAAGCGAATTTTCAAGAGGCATATGAGAAAGGGATCCATAAGTCAAAGTTTGGGAGCA 510

Db 303 ACATATGAGGATTCCTCAACCTGATTTGCTGCTGTTTNCCTGTTGCTAGCTGCAATAATGT 362

Qy 511 ACATATGAGGACTGCTTGTAGTTGATT-GCTCAAGTTT-CCTGTTGTTG-CAGCTTTA-TGT 566

Db 363 TTATCCGAGGAGTATATAANAAT 386

Qy 567 TTATCCGA--GGATGTATAAGAAT 588

RESULT 8

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

```

KEYWORDS EST
SOURCE domestic silkworm.
ORGANISM Bombyx mori

REFERENCE
1 (bases 1 to 620)
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Lepidoptera; Bombycoidea; Bombycidae; Bombyx.
AUTHORS Mita K., Morimyo M., Shimada T., Okano K. and Maeda S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT On Sep 1, 1995 this sequence version replaced.

FEATURES
source
1..620
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsu@nirs.go.jp
PROJECT - "CREST project by JST".
Location/Qualifiers
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="n0357"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT 162 a 141 c 146 g 171 t
ORIGIN

Query Match 11.5%; Score 178; DB 26; Length 620;
Best Local Similarity 65.3%; Pred. No. 8.20e-108;
Matches 380; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Db 1 GACCACGAGGGTGAACAGCTTTGTCGCACACTACGCACTTTAGTGGTCTCTCTTAGT 60
QY 706 GATCATGAGGTGGAAATTTAGTGCACACACTGGCATTTGGTGGTASTCCATTCA 755
Db 61 GATCCTTACTTATCTTGGCGGTGGACTCAATGGTCTTCTGGAGCACCTTACGGCGTG 120
QY 766 GATCCTTATTTGTCATTTGAGAGAGCAATTAATGGTTGGCTGGGCACTCCATGATTA 825
Db 121 GCTAATCAAGAGTTCCTGGTATGGCTCGAGAACTGGCAAACTGGCAAACTGGTATATTC 180
QY 826 CCCAACGAGGAGTCTGCTGTGGATTAATCAATGTTGTGATGAATGTGGAGAGACATC 885
Db 181 ACAGAAGAGCAACTCAAGAAATTCATCTGGAACACACTTAAATCTGGTCAGGTGTACCT 240
QY 886 TCGACAGCAGCTTGAAGATTATGTTTGAAGACACTAAACAGTGGCAAGGTGTACCT 945
Db 241 GGTACGGTCATGAGTACTTAGNAACTGATCAAGATATACCTTGCAGCGTGAGTTT 300
QY 946 GGATTTGGTCTAGAGTATTGGGGAAGACAGATCCAAAGATACACATGCCAAAGAAATTT 1005
Db 301 GCTCTTAAGCACTTACCAATGACCCATTATCAAGTGTGCTGCTCTTTTACAAGTTT 360
QY 1006 CGGTGGAAGCACTTGCCTGATGACCCATTTTTCATTTGGTGTCAAGTGTGTAAGTG 1055
Db 361 GTTCGGGTGATCTCACTGCAACTTGGCAAGTCAAGAACCCATGGCTTAATGTAGACTCC 420
QY 1066 GTGCTCTCTATTCTATTAGAGCTTGAAGAGTAAAGATCCATGGCTTAATGTGATGCT 1125
Db 421 CATTCGGGAGTTCCTTTGTCAGTATTATGGTCTGAAGAGATGACTACTATACAGTAATG 480
QY 1126 CATAGTGAGTTCCTTGAACCACTATGGTTTGACAGAGCAAGATACATACGCTTTTG 1185
Db 481 TTTGAGTGTCCCGAGGCTTGGGTGTTCTCGCTCAGTGTGATTTGGTCCCGTGGCTCGGA 540
QY 1186 TTTGGGTATCAAGAGTCTTGAATATGTCACAGCTATATGGGACCGAGCTCTTGGC 1245
Db 541 CTTCCTTGTAGAGCAAAAGTCTGCTCAGCAGCGAGATGCTT 582
QY 1246 TTGCCGTAGAGGCCAAAGAGTGTCTATGGAATGGCTT 1287

RESULT 9

```

```

LOCUS AA313713 501 bp mRNA EST 19-APR-1997
DEFINITION EST185581 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end
similar to similar to citrate synthase, mRNA sequence.
ACCESSION AA313713
NID G1966042
VERSION AA313713.1 GI:1966042
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geoghagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald J.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Saudek D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrie A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei Y.F., Wang J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseitine W.A., Fields C.,
Fraser C.M. and Venter J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Apr 14, 1993 this sequence version replaced gi:693326.
Other ESTs: THCL171646
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M3 Reverse

FEATURES
source
Location/Qualifiers
1..501
/organism="Homo sapiens"
/note="Organ. colon. Vector. pBluescript SK-, Site_i.
EcoRI; Site_2: XhoI"
/db_xref="ATCC (inhost):109839"
/db_xref="taxon:9606"
/map="11"
/clone_lib="Colon carcinoma (HCC) cell line"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon
carcinoma;Dukes B2"
BASE COUNT 128 a 117 c 142 g 114 t
ORIGIN

Query Match 10.6%; Score 164; DB 11; Length 501;
Best Local Similarity 66.9%; Pred. No. 7.09e-97;
Matches 324; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Db 1 TGCAGAGCCATGAACGGCTGGCAGGCGCTCTCCATGGACTGGCAATCAGGAAGTCT 60
QY 783 TGCAGCAGCATTAATGATGGTTTGGCTGGCGCCCTCCATGGACTAGCCACCAAGAGTCT 842
Db 61 TGTCTGGCTAACACAGCTGTCAGAGGAAGTGTGCAAGAGATGTGTGATGAGAACTTACG 120
QY 843 GCTGTGGAATTAATCAGTTGTTTCATGAATGTGGAGAGACATCTCGACAGACAGTTGAA 902

```

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo.

/db_xref="taxon:10090"
/map="3"
/clone="IMAGE:679873"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"

BASE COUNT 129 a 125 c 150 g 118 t

ORIGIN

Query Match 10.1%; Score 156; DB 10; Length 522;

Best Local Similarity 66.4%; Pred. No. 1.17e-90;

Matches 346; Conservative 0; Mismatches 172; Indels 3; Gaps 3;

Db 1 ATGAGGGTGTAAATGTAAGTGGCCACACAGCCATTTGGTGGG-AGCGCTGTT-CAGACC 58

Qy 710 ATGAGGGTGAATATGTTAGTCACACACTGCGCATTTGGTGGGAGTCCACATTCAGATC 769

Db 59 CTTACCTGTCTCTTTCAGCAGCCATGAATGGCTGGCGGGCT-CTACATGGACTAGCAA 117

Qy 770 CTTATTTGTCATTTGCAGCAGCATTAATGGTTTGGCTGGCCACTCCATGGATTAGCCA 829

Db 118 ATCAGGAGGTGCTTGTCTGTGCTGACACAGCTACAGAGGAAGTTGGCAAGACGTGTCTAG 177

Qy 830 ACCAGGAAGTCTGCTGTGGATTAAATCAGTTGTTGATGATGAGAGAACATCTCGA 889

Db 178 ATGAGAAGTTACGAGACTACATCTGGAACACACTCAATTCAGGACGGTGGTCCAGAT 237

Qy 860 CAGAGCAGTTGAAAGATTATGTTTGGAGACACTAAACACAGTGGCAAGGTTGTACTGGAT 949

Db 238 ACGGTGTCAGTACTGAGGAAGACTGACCCCTGCTATCTCTGTCAGGAGAGTTTGTCT 297

Qy 950 TTGGTCTAGGAGTATTCGGGAAGACAGATCCAAGATACACATGCAAGAGAAATTCGGT 1009

Db 298 TGAACATCTGCTTAAGATCCCATGTTTCAAGCTGGTGGCTCAGCTGTACAGATTTGTC 357

Qy 1010 TGAAGCACTTGCCTGATGACCCATTTTCAATTGGTGTCAAGTTGTATGAAGTGTGC 1069

Db 358 CCAATATCTCTTAGCAAGGAAGGCTAGAACCCCTTGCCCAAGCTAGACCTCACA 417

Qy 1070 CTCCTATCTTATAGAGCTTGGAAAGGTAAGAAATCCATGCCCTAATGTGATGCTCATA 1129

Db 418 GTGGGGTCTCTCCAGTACTATGCGATGAGGAGATGAAGTACTACACAGTCTCTGTTG 477

Qy 1130 GTGGAGTTTCTCTGAACCACTATGTTTGACAGAAGCAAGATACATACGSGTTTGTG 1189

Db 478 GAGTGTCTGGGCACTGGGTGTGCTAGCCCAAGCTCATCTGG 518

Qy 1190 GGGTATCAAGGAGTCTTGGAAATATGCTCACAGCTTATATGG 1230

RESULT 11

LOCUS AA231705 307 bp mRNA EST 27-FEB-1997

DEFINITION CDO534.R cDNA from oat *Avena sativa* cDNA clone CDO534, mRNA

sequence.

ACCESSION AA231705

NID g1854091

VERSION AA231705.1 GI:1854091

KEYWORDS EST.

SOURCE oat.

ORGANISM *Avena sativa*

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; *Avena*.

REFERENCE 1 (bases 1 to 307)

AUTHORS Vandeynze,A.E., Sorrells,M.E., Park,W.D., Ayres,N.M., Fu,H.,

Carlinhour,S.W. and McCouch,S.R.

TITLE Anchor Probes for Comparative Mapping of Grass Genera

JOURNAL Unpublished (1997)

COMMENT On Nov 29, 1993 this sequence version replaced gi:634857.

Db 121 AGACTACATCTGAAACACACTCAACTCAGCAGGGTGTTCAGAGCTATGGCCATCGACT 180

Qy 903 AGATTATGTTGGAACACTAAACAGTGGCAAGTTGTACTGATTTGGTCTAGGAGT 962

Db 181 ACTAAGAAGACTGATCCGGCATATACCTGTGACGAGAGTTGTCTGGAACACCTGCC 240

Qy 963 ATTGCGGAACAGATCCCAAGATACACATGCCAAGAGAATTTGGTTGAAGCATGGC 1022

Db 241 TAATGACCCCATGTTTAAGTGTGCTCAGCTGTACAGATTTGCCCCAATGCTCTTT 300

Qy 1023 TGAATGACCCATTTTCAATTTGGTGTCAAGTTGTATGAAGTGGTGCCTCTTCTATT 1082

Db 301 AGACGAGGTAAAGCAAGTAATCTTGGCCCAATGTAGATGTCTACAGTGGGTGCTGCT 360

Qy 1083 AGAGCTTGAAGAGTAAGCAATCCATGGCTTAATGTGATGCTCATAGTGGAGTTTGT 1142

Db 361 CCAGTATTATGATGACGAGGAGATGAATTAATACACAGGTCCTGTTGGGTGTACAGGAC 420

Qy 1143 GAACCACTATGTTTTCAGCAAGCAAGATACTATACGGTTTGTGTTGGGTATCAAGGAS 1202

Db 421 ATTGGGTGTACTGGCAGCTCATCTGAGCGAGCGCTTAGGTTCCCTCTAGAAAGGC 480

Qy 1203 TCTTGGNATATGCTCACAGCTTAATATGGACCGAGCTTGGCTTCCCGCTAGAGAGGCC 1262

Db 481 CAAG 484

Qy 1263 AAAG 1266

RESULT 10

LOCUS AA237580 522 bp mpNA EST 03-MAP-1987

DEFINITION mx11401.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:679873 5'

similar to SW:Clisy.PIG P00889 CITRATE SYNTHASE, MITOCHONDRIAL

PRECUSOR ., mRNA sequence.

ACCESSION AA237580

NID g1861602

VERSION AA237580.1 GI:1861602

KEYWORDS EST.

SOURCE house mouse.

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 522)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1392879.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:419577

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 453.

Location/Qualifiers

1. .522

/organism="Mus musculus"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACATCTGAAGTGGGAGCGCGGAATCTTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors

FEATURES

Source

Contact: McCouch SR
Dept Plant Breeding
Cornell University
Ithaca, NY 14853-1901, USA
Tel: 607 255 0420
Fax: 607 255 6683
Email: srm4@cornell.edu
cDNA from oat (*Avena sativa*); reverse sequence of RFLP probe
CDO534. Sequence determined by Nicola M. Ayres.

FEATURES

source

1..307
Location/Qualifiers
/organism="Avena sativa"
/cultivar="Brooks"
/note="Vector: Uni-ZAP XR/pBluescript; Site_1: EcoRI;
Site_2: XhoI; A Uni-ZAP XR cDNA library was constructed
from etiolated leaf mRNA from the oat cultivar 'Brooks'
and converted to pBluescript (amp resistant) as described
in Heun et al. (1991) Genome 34:437-447. For insert
amplification, use M13 forward and reverse primers.
Clones from this library are designated with the prefix
'CDO'. *Note: Clone CDO1061 was recloned into the TA
cloning vector and carries kanamycin resistance."
/db_xref="taxon:4498"
/clone="CDO534"
/clone_lib="cDNA from oat"
/tissue_type="etiolated leaf"
BASE COUNT 84 a 64 c 79 g 80 t
ORIGIN

Query Match 9.7%; Score 151; DB 10; Length 307;
Best Local Similarity 75.8%; Pred. No. 8.76e-87;
Matches 222; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Db 1 CGGCACTGAACGGTTAGCTGGCCACATGCATGGCTGGCTAATCAGGAAGTGTGCTA 60
QY 787 GCAGCAATTAATGTGGCTGGCCACATCCATGGATTAGCCACAGGAAGTCTCTGTG 846
Db 61 TGGATCAATCTCTGATGGAAGAAACCGGGAGTAAACATACAACTGATCAGCTTAAAGAA 120
QY 847 TGGATTAAATCAGTTCTGATGAATGGAGAGACAATCTGCAGAGACATTTGAAGAT 906
Db 121 TAGCTATGGAAGACACTGAAGAGTGAAGAGTTGTTCTCGGTATGGTCACTGGAGTTCTA 180
QY 907 TATGTTTGGGAAGACATAAACAGTGGCAAGTTGTACCTGGATTGGTCTAGGAGTATTG 966
Db 181 CGTAATACAGATCCAGTACTCGTCCAAAGGGAGTTGCACTTAAGTATTACCCGAG 240
QY 967 CGGAAGACAGATCCAGATACATGCCAAAGAGAAATTTGCGTTGAAGCACTTGCCTGAT 1026
Db 241 GACCCACTTTCCAACTGCTCCAAAGTTGTACGAAGTTGTGCCCTCTATCCT 293
QY 1027 GACCAATTTTCAATTTGGTCAAGTTGTATGAAGTGTGCTCTCTATCT 1079

RESULT 12
LOCUS AAL12941 466 bp mRNA EST 01-AUG-1997
DEFINITION zn59b06.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
IMAGE:562451 5' similar to SW:CI5V.PIG P00889 CITRATE SYNTHASE,
MITOCHONDRIAL PRECURSOR ;, mRNA sequence.

ACCESSION AAL12941
NID gl664291
VERSION AAL12941.1 GI:1664291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 466)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE
JOURNAL
COMMENT

WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:693136.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

This clone is available royalty-free through LILN: contact the
IMAGE Consortium (info@image.liln.gov) for further information.

Insert Length: 2292 Std Error: 0.00
Seq Primer: -28M13 rev2 from Amersham
High quality sequence stop: 423.

FEATURES

source

Location/Qualifiers
1..466
/organism="Homo sapiens"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Skeletal muscle from patient with
malignant hyperthermia. Average insert size: 1.0 Kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'."
/db_xref="GDB:4595741"
/db_xref="taxon:9606"
/clone="IMAGE:562451"
/clone_lib="Stratagene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT 117 a 116 c 122 g 110 t 1 others
ORIGIN

Query Match 9.2%; Score 143; DB 36; Length 466;
Best Local Similarity 68.8%; Pred. No. 1.31e-80;
Matches 287; Conservative 0; Mismatches 127; Indels 3; Gaps 3;

Db 50 AATTTCCACCAACATGTTAGGCTATACATCATCAGTTCACTGAGCTCACGCGCTGTAC 109
QY 631 AATTTCCACACATGTTGGGATTTGATAGCCCTCAGATGCTTGAGCTGATCGCCTTAT 690
Db 110 CTCACCATCCACAGTGCACCATGAGGTTGGCAATGTAAGTATGCCATACACAGCCATTTGGT 169
QY 691 GTTCACAATTCACAGTGATCATGAGGTTGAAATGTTAGTGACACACTGGCATTTGGTG 750
Db 170 GGCAGTGGCTNTTCCGACCCGTTACTCTCTTTCACAGCCCATGACGGCTGGCAGG 229
QY 751 GGTAGTCCACTTTCAGATCC-TTATTTGTCAATTTGCAGCAGCATTAATGTTGGCTGG 809
Db 230 GGT-CTCCATGACCTGGCAAAATCAGAAAGTCTGTCTGCTTAACACAGCTGCAGAAAGGA 288
QY 810 GGCACCTCCATGATAGCCACACAGCAAGTCTGCTGTGGATTAAATCAGTTGTTGATGA 869
Db 289 AGTTGGCAAGATGTTGTCAGATGAGAAGTTAGGAGCTACATCTGGACACACATCAACTC 348
QY 870 ATGTGGAGAGAACATCTCGACAGACAGCTTGAAGATTATGTTTGGAGACACTTAAACAG 929
Db 349 AGCAGCGGTTGTCAGGCTATGGCAT-CCAGTACTAAGGAAGACTGATCCGGCATATAC 407
QY 930 TGGCAAGTTTACCTGGATTGCTAGGATTTGGCTAGGAGTATTCGGGAAGACAGATCAACATAC 989
Db 408 CTGTACGACGAGAGTTTGGCTCTCAAAACACCTGCCTCAATGACCCCATGTTTAAGTTGGT 464
QY 990 ATGCCAAAGAGAAATTTGGCTTGAAGCACTTGCCTGATGACCCATTTTTCATTTGGT 1046

RESULT 13
LOCUS AA498291 473 bp mRNA EST
DEFINITION vh38g11.r1 Barstead mouse pooled organs MFLRB4 Mus musculus cDNA
clone IMAGE:889316 5' similar to SW:CI5V.PIG P00889 CITRATE
SYNTHASE, MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION AA498291

QY	900	GAAGATTATGTTTGAAGACACTAAACAGTGGCAAGGTTTACTCGATTGGTCTAGG	959
DB	321	AGTACTGAGGAAGACTGACCCCTCGCTATTCTGTCTAGAGAGAGTTTGTCTTGAACATCT	380
QY	960	AGTATTGCGGAAGACAGATCCAAAGATACACATGCCAAAGAGAAATTTGGCTTCAAGCACTT	1019
DB	381	GCTTAAGATGCCATGTTTCAAGCTGGTGGCTCAGCTGTACAAAGATTGTGGCCCAATATCCT	440
QY	1020	GCTGTATGACCCATTTTTCATTTGGTGTCAAAGTTGTATGAAGTGGTGGCTCTCTATCT	1079
DB	441	CTTAGAGCACGGGAAGGCTTAAGAACCCCTTGGCC	473
QY	1080	ATTAGAGCTTGGAAAGGTAAGAATCCATGCC	1112
RESULT	14		
LOCUS	217455	213 bp	mRNA
DEFINITION	ATTS0029 AC16H Arabidopsis thaliana cDNA clone TAT2B1 5' similar to		10-NOV-1992
SOURCE	CITRATE SYNTHASE, MITOCHONDRIAL PPRECURSOP. Swiss-Prot entry P00889,		
ACCESSION	217455		mRNA sequence.
NID	916597		
VERSION	217455.1	GI:16597	
KEYWORDS	EST.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1 (bases 1 to 213)		
AUTHORS	CNRS.		
TITLE	The Arabidopsis thaliana transcribed genome: the GDR cDNA program		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Bardet C., Axelos M., Tremousaygue D., Lebas M., Lagraverie T., Lescure B.		
FEATURES	1..213		
source	/organism="Arabidopsis thaliana"		
	/strain="ecotype Columbia"		
	/note="Vector: Lambda ZAPII; tissue=cell suspension		
	culture of ecotype Columbia; clone.library=AC16H; Cloning		
	vector: Lambda ZAPII; Physiological condition: cycling		
	cells."		
	/db_xref="taxon:3702"		
	/clone="TAT2B1"		
	/clone_lib="AC16H"		
BASE COUNT	59 a 40 c 52 g	62 t	
ORIGIN			
Query Match	8.7%	Score 135; DB 8; Length 213;	
Best Local Similarity	84.4%	Pred. No. 1.83e-74;	
Matches	178; Conservative	0; Mismatches 31; Indels	2; Gaps 2;
DB	5	TCACCATCCACAGTATCATGAAGTGGAAATGTTAGTGCTCACACTGGTACCTGGTTG	64
QY	692	TCACAATTCACATGATCATGAGGTGGAAATGTTAGTGCACACACTGGCCATTTGGTGG	751
DB	65	GTAGTCACATTTTCAGACCCATATCTGTCATTTTGCAGCTGCATTAATGGTTAGCTGGC	124
QY	752	GTAGTCACATTTTCAGATCCTTTATTTGTCATTTTGCAGCAGCAATTAATGGTTGGCTGGC	811
DB	125	CATCTCATGTTTGGCTAATCAGGAAGTTTTCCTTTGGATCAA-TCAGTCGTAGAGGAAT	183
QY	812	CATCTCATGATTAGCAACACAGGAGTCTCTGTGGATTAATCATGTTTGTATGAAT	871
DB	184	GTGAGAGATATATCAA-AGAACAGTTGAA	213
QY	872	GTGAGAGACATCTTCGACAGCAGTTGAA	902

WQRELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 15:45:02 1999; MasPar time 17.45 Seconds
Tabular output not generated. 532.612 Million cell updates/sec

Title: US-08-702-718-4
Description: (1-437) from US08702718.pep
Perfect Score: 3243

Sequence: 1 SSNLDLRFSELOFLPEQQER LERPKSVTMEWLEKFKRRA 437

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 35.586; Variance 154.962; scale 0.230

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	3243	100.0	437	14	Sugar beet citrate sy	3.53e-302
2	2800	86.3	469	14	Tobacco citrate synth	1.87e-258
3	2637	81.3	471	15	Potato citrate synth	2.20e-242
4	2637	81.3	471	14	Potato citrate synth	2.20e-242
5	312	9.6	436	3	Citrate synthase enco	4.10e-17
6	103	3.2	657	23	Methionyl-tRNA synth	1.28e-01
7	101	3.1	456	34	Methyl tetrahydropter	1.77e-01
8	99	3.1	566	39	An antitumour protein	2.45e-01
9	100	3.1	738	31	New DNA sequence isol	2.08e-01
10	93	2.9	219	17	AFT-1 interacting pro	6.36e-01
11	94	2.9	237	21	H. pylori cell envelo	5.44e-01
12	93	2.9	324	15	Equine herpesvirus (E	6.36e-01
13	93	2.9	324	9	Equine herpesvirus US	6.36e-01
14	95	2.9	329	29	H. pylori ORF 06p302	4.64e-01
15	95	2.9	329	29	H. pylori ORF hp3el10	4.64e-01
16	94	2.9	461	22	H. pylori cell envelo	5.44e-01

17 93 2.9 471 14 R79163 Partial sequence of b 6.36e+01
18 93 2.9 552 35 W1636 Omega-cyclohexane fat 6.36e+01
19 93 2.9 754 8 R45000 Cellulose synthase op 6.36e+01
20 93 2.8 254 32 W3688 Human secreted protei 1.02e+02
21 90 2.8 351 37 W74904 Human secreted protei 1.02e+02
22 90 2.8 355 24 W29179 Rat CC chemokine rece 1.02e+02
23 92 2.8 430 36 W1618 Staphylococcus aureus 7.44e+01
24 92 2.8 430 33 W24482 Class II EPSPs used f 7.44e+01
25 92 2.8 430 25 W36887 Class II EPSP synthas 7.44e+01
26 90 2.8 480 17 R55310 Cystathionine gamma s 1.02e+02
27 90 2.8 509 17 R55311 Cystathionine gamma s 1.02e+02
28 91 2.8 564 20 W01464 NHI Hxub protein. 8.70e+01
29 91 2.8 565 20 W01463 Lactobacillus amylovo 8.70e+01
30 91 2.8 720 23 W19266 Peptide derived from 1.88e+02
31 86 2.7 58 39 W87538 Chimeric anti HIV ant 1.88e+02
32 86 2.7 427 10 R0303 EPSP synthase from E. 1.88e+02
33 89 2.7 456 17 R04400 Tyrosine phenol-lyase 1.19e+02
34 89 2.7 456 17 R0994 Tyrosine phenol-lyase 1.19e+02
35 87 2.7 501 5 R27741 Sequence transcribed 1.61e+02
36 87 2.7 547 35 W1526 Helicobacter polysept 1.88e+02
37 86 2.7 556 17 R96737 A. niger Bo-1 carboxy 1.88e+02
38 89 2.7 668 26 W38164 Pmel17 encoded by cdv 1.19e+02
39 89 2.7 831 32 W34372 Carboxydothexmus hydr 1.88e+02
40 86 2.7 968 28 W41310 CF-5 pathogen resista 1.61e+02
41 87 2.7 968 28 W41309 CF-5 pathogen resista 1.61e+02
42 87 2.7 1016 28 W41312 CF-5 pathogen resista 1.61e+02
43 88 2.7 1195 39 W75420 T. thermophilus nitrat 1.38e+02
44 88 2.7 4472 19 R97245 Virulence gene cluste 1.19e+02
45 89 2.7 4472 19 R97245

ALIGNMENTS

RESULT 1
ID R82839 standard; Protein; 437 AA.
AC R82839:
DT 25-JAN-1996 (first entry)
DE Sugar beet citrate synthase.
KW Citrate synthase; flower formation.
OS Beta vulgaris strain Zuchtlinie 5S 0026
PN W09524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR N-PSDB; T04200.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Disclosure: Page 57-60; 87pp; English.
CC To identify a cDNA from sugar beet which codes for citrate
CC synthase, a cDNA bank of leaf tissue from sugar beet was prepd.
CC Plaques of this cDNA bank were screened using radioactive DNA
CC probes which comprise a mixture of Solanum tuberosum citrate
CC synthase cDNA (T04199) and Nicotian tabacum citrate synthase cDNA
CC (see T04201). One of the clones was sequenced. The nt. sequence is
CC given in T04200.
SQ Sequence 437 AA:
Query Match 100.0%; Score 3243; DB 14; Length 437;
Best Local Similarity 100.0%; Pred. No. 3.53e-302;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ssndlrslqlpeqqrkkkkgfsgfqlgnvdmvlgmrgmtglwetslldp 60
QY 1 SSNLDLRFSELOFLPEQQERLKKIKKEFGSLQGNINVDVGLGMRGMTGLWETSLDP 60
Db 61 esqirfrfsspecokllpaasagaepleglllllltqkvskevvdalsadlrkrsi 120

QY 61 EEGIRPGFSIPESCKLLPAASAGAEPLPEGLLWLLLTGKVPSEQVDALSADLRKASI 120
 Db 121 pdhvyktidalphtmqfctgvmalqtrsefkyakeghkshkfwepdyedclsia 180
 QY 121 PDHWYKTIDALPITAHPTMTQFCTGVNALQTRSEFQKAYEKGIIHKSFWEPDYEDCLSIA 180
 Db 181 qvvaayvyrmykqvgplddslvgantfahmlgfdsgmlelmrllyvthsdhegg 240
 QY 181 QVPVAAAYVYRRMYKNGQVPIPLDSDLYGNGFAHMLGFDSPQMLELMRLYVTHSDHEGG 240
 Db 241 nvsahgtghlvgspldspylsfaaalinglagplhglanqevllwiksvvdecgenisteql 300
 QY 241 NVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENISTEQL 300
 Db 301 kdyvwtlmsgkvvpvgfagvlrktprytcqrefalkhlpdpffqlvsklyevvppil 360
 QY 301 KDYVWTKLNSGKVVPVGGVLGRKTDPRYTCQREFALKHLPDPFFQLVSKLYEVVPPIL 360
 Db 361 lelqkvknpwpnvahsgvllnhhygltearyyvtlfgvrsalgicsqliwdralqlpieler 420
 QY 361 LEIGKVKNPWPNDVDAHSGVLLNHYGLTEARYYVTLFGVRSRSLGICSQLIWDRALGLPLER 420
 Db 421 pksvtmewlekckrra 437
 QY 421 PKSVTMEWLEKCKRRA 437

RESULT 2

ID R82840 standard; Protein: 469 AA.
 AC R82840;
 DT 25-JAN-1996 (first entry)
 DE Tobacco citrate synthase.
 KW Citrate synthase; flower formation.
 OS Nicotiana glauca
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 DR WPI: 95-328278/42.
 DR N-PSDB: T04201.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 to improve storage of tubers, etc. and to reduce sprouting
 PS Disclosure: Page 60-63; 87pp: English.
 CC To identify a cDNA bank of leaf tissue from tobacco for citrate
 synthase, a cDNA bank of leaf tissue from tobacco was prep'd.
 CC Plaques of this cDNA bank were screened using a radioactive DNA
 probe which comprises Solanum tuberosum citrate synthase cDNA
 CC (T04199). One of the clones was sequenced. The nt. sequence is
 CC given in T04201.
 SQ Sequence 469 AA;

Query Match 86.3%; Score 2800; DB 14; Length 469;
 Best Local Similarity 85.7%; Pred. No. 1.87e-258;
 Matches 373; Conservative 34; Mismatches 27; Indels 1; Gaps 1;

Db 35 ssqldlrselqelipeqqdrllkiksehkvgqlgnitvdmvlgmrgmtgllwetslldp 94
 QY 1 SSNLDLRSLELQELIPEQQERLKKIKKEFGSFGQGNINVDVLMGGMRGMTGLLWETSLLDP 60
 Db 95 degirirgisieqcqvlpaaakpggeplpegllwllltgkvpskeqvdslsqratsv 154
 QY 61 EEGIRPGFSIPESCKLLPAASAGAEPLPEGLLWLLLTGKVPSEQVDALSADLRKASI 120
 Db 155 pdhvyktidalphtmqfctgvmalqtrsefkyakeghkshkfwepdyedclsia 214
 QY 121 PDHWYKTIDALPITAHPTMTQFCTGVNALQTRSEFQKAYEKGIIHKSFWEPDYEDCLSIA 180
 Db 215 qvplvaayvyrmykngntipkddslvgantfahmlgfdsgmlelmrllyvthsdhegg 274

QY 181 QVPVAAAYVYRRMYKNGQVPIPLDSDLYGNGFAHMLGFDSPQMLELMRLYVTHSDHEGG 240
 Db 275 nvsahgtghlvasalsdpylsfaaalinglagplhglanqevllwiksvveecgeniskeql 334
 QY 241 NVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENISTEQL 300
 Db 335 kdyawklkskvvpvgfghvrlrktprytcqrefalkhlpdpffqlvsklyevvflql 394
 QY 301 KDYVWTKLNSGKVVPVGGVLGRKTDPRYTCQREFALKHLPDPFFQLVSKLYEVVPPIL 360
 Db 395 qnlaki-npwpnvahsgvllnhhygltearyyvtlfgvrsalgicsqliwdralqlpieler 453
 QY 361 LEIGKVKNPWPNDVDAHSGVLLNHYGLTEARYYVTLFGVRSRSLGICSQLIWDRALGLPLER 420
 Db 454 pksvtmewlenhckk 468
 QY 421 PKSVTMEWLEKCKR 435

RESULT 3

ID R86383 standard; Protein: 471 AA.
 AC R86383;
 DT 24-APR-1996 (first entry)
 DE Potato citrate synthase.
 KW Citrate synthase; inhibitor; increased storage capacity; potato;
 KW antisense DNA.
 OS Solanum tuberosum.
 PN DE4408629-A1.
 PD 14-SEP-1995.
 PF 09-MAR-1994; 408629.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-Roeber B, Landschuetze V;
 DR WPI: 95-321536/42.
 DR N-PSDB: T03410.
 PT Inhibiting citrate synthase (CS) activity in plants - to inhibit
 flower formation and improve storage capacity, e.g. in potatoes,
 also new CS sense and anti-sense DNA sequences
 PS Claim 5; Page 15-19; 35pp: German.
 CC The potato citrate synthase (CS) gene or cDNA sequence encoding CS
 CC (T03410) may be used to produce antisense CS sequences. CS DNA
 CC sequences are useful for altering CS activity in plants. Antisense
 CC CS sequences can be used to inhibit CS expression in plants and has
 CC the effect of inhibiting flower formation and by doing so improves
 CC the plant's storage capacity. This is partic. useful in crop plants
 CC of any kind but esp. useful in potatoes. In addition to altering CS
 CC activity the DNA sequences can also be used to identify similar
 CC transgenic plants with altered CS activity. This sequence
 CC represents the amino acid sequence of potato citrate synthase.
 SQ Sequence 471 AA;

Query Match 81.3%; Score 2637; DB 15; Length 471;
 Best Local Similarity 81.3%; Pred. No. 2.20e-242;
 Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;

Db 35 ssqldlrselqelipeqqdrllkiksehkvgqlgnitvdmvlgmrgmtgllwetslldp 93
 QY 1 SSNLDLRSLELQELIPEQQERLKKIKKEFGSFGQGNINVDVLMGGMRGMTGLLWETSLLD 59
 Db 94 degirirgisieqcqvlpaaakpggeplpegllwllltgkvpskeqvdslsqratsv 153
 QY 60 EEGIRPGFSIPESCKLLPAASAGAEPLPEGLLWLLLTGKVPSEQVDALSADLRKAS 119
 Db 154 lsliimtytidalpvtahptmqfctgvmalqtrsefkyakeghkshkfwepdyedclsml 213
 QY 120 IPDHV-YKTIDALPITAHPTMTQFCTGVNALQTRSEFQKAYEKGIIHKSFWEPDYEDCLS 178
 Db 214 laqvplvaayvyrmykngntipkddslvgantfahmlgfdsgmlelmrllyvthsdh 273

QY 179 IAQVPVAAAYVPRMYKNGQVLPDSDLDYGGNFAHMLGFDSPQMLEL-MRLVVTIHSOH 237
 Db 274 eggnvsahtghlvasalsdpyisfaaalnglqplhglanqevlllwiksvveecgenisk 333
 QY 238 EGNVSAHTGHLVGSPLSDPYISFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENIST 297
 Db 334 eqlkdyvwtlinsgkvvgfghvrlrktvprytcqrefamkhlpedplfqlvsklyevf 393
 QY 298 EQLKDYVWTKLNSGKVVPFGGLGVLKRTDPRYTCQREFALKHLPDDPPFQLVSKLYEVP 357
 Db 394 lfqlnklakl-pwpsvdahsgvllnyygltearyyvtlfgvsralgicsqlldwralgpl 452
 QY 358 PILLELGKVKNPWPNVDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLP 417
 Db 453 lerpkvmtewlenqckk 470
 QY 418 LERPKSVMTEWLEKFCR 435

RESULT 4

ID R82838 standard; Protein; 471 AA.
 AC R82838;
 DT 25-JAN-1996 (first entry)
 DE Potato citrate synthase.
 KW Citrate synthase; flower formation; tuber storage.
 OS Solanum tuberosum.
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR N-PSDB; T04199.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Disclosure; Page 53-56; 87pp; English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of
 CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. a cDNA library was prepd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 471 AA;

Query Match 81.3%; Score 2637; DB 14; Length 471;
 Best Local Similarity 81.3%; Pred. No. 2,20e-242;
 Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;
 Db 35 ssgldrlselvqelipeqqdrlkkidsdkg-signitdmvlgmrgmtglwxphyid 93
 QY 1 SSSNLDLSEL-QELIPEQERLAKIKKEFGSFLQNLGNVDMVLGMRGTMGLWETSLD 59
 Db 94 pdeqirfirlsiecckvlpaaqpggeplpegillllltokvpskeqvasivsgiaesgi 153
 QY 60 PEEGIRFGFSIPPCQKLLPAASAGAEPLPEGLLWLLTGTGKVPKQVDSALSLKRRAS 119
 Db 154 isliimvttidalpvtahmtqtatgymalqvqsefqqkayekghsksyweptyedsml 213
 QY 120 IPDHV-VKTDALPITAHPTQCTGVMALQTRSEFQKAYEKGIHKSKEWPEYEDCLSL 178
 Db 214 iaqvpvvaayvrrmykngdtipkdesldyganfahmlgfssemhellmrlvytihsdh 273
 QY 179 IAQVPVAAAYVPRMYKNGQVLPDSDLDYGGNFAHMLGFDSPQMLEL-MRLVVTIHSOH 237
 Db 274 eggnvsahtghlvasalsdpyisfaaalnglqplhglanqevlllwiksvveecgenisk 333

QY 238 EGNVSAHTGHLVGSPLSDPYISFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENIST 297
 Db 334 eqlkdyvwtlinsgkvvgfghvrlrktvprytcqrefamkhlpedplfqlvsklyevf 393
 QY 298 EQLKDYVWTKLNSGKVVPFGGLGVLKRTDPRYTCQREFALKHLPDDPPFQLVSKLYEVP 357
 Db 394 lfqlnklakl-pwpsvdahsgvllnyygltearyyvtlfgvsralgicsqlldwralgpl 452
 QY 358 PILLELGKVKNPWPNVDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLP 417
 Db 453 lerpkvmtewlenqckk 470
 QY 418 LERPKSVMTEWLEKFCR 435

RESULT 5

ID R14357 standard; Protein; 436 AA.
 AC R14357;
 DT 23-JAN-1992 (first entry)
 DE Citrate synthase encoded by aarA.
 KW Acetic acid resistance.
 OS Acetobacter sp.
 PN J03219878-A.
 PD 27-SEP-1991.
 PF 05-FEB-1990; 024395.
 PR 15-FEB-1989; JP-033776.
 PR 05-FEB-1990; JP-024395.
 PA (NAKA-) NAKANO SUMISE KK.
 DR WPI; 91-329112/45.
 DR N-PSDB; Q14367.
 PT Acetic acid resistant gene with in plasmid and transformed
 PT acetobacter - improves yield of acetic acid fermentation.
 PS Disclosure; Fig 5; 12pp; Japanese.
 CC The aarA gene encodes the citrate synthase and is part of an
 CC acetic acid resistance operon comprising aarA, aarB and aarC.
 CC The DNA can be used to prepare an acetic acid resistant strain of
 CC bacteria for use in acetic acid fermentation.
 CC See also R14358 and 59.
 SQ Sequence 436 AA;

Query Match 9.6%; Score 312; DB 3; Length 436;
 Best Local Similarity 31.1%; Pred. No. 4.10e-17;
 Matches 102; Conservative 85; Mismatches 108; Indels 33; Gaps 29;
 Db 91 eeviyllingelpnkaqydtfntltntnlhheqirfnfgrddahpmailegtvgals 150
 QY 90 EGLLWLLTGTGKVPKQVDSALSLKRRASIPDHVYKTIDALPITAHPTQCTGVMALQ 149
 Db 151 a-fypdandiap-an-rdia--a-mrliaiptiaawayk--ytgqeaifiyrnd-ln 200
 QY 150 TRSEFQKAYEKGIHKSKEWPEYEDCLSLIAQVPVAAAYVPRMYKNGQV-I-PLDSDLD 207
 Db 201 yaenflsmfmfarmsepkyknpvlaramrillilbadheq-naststvtlragstganpfac 259
 QY 208 YGGNFAHML-G-FDSPQMLE--LMR-LY-VTI-HSDHEGGNVSAHTGHLVGSPLSDPYLS 260
 Db 260 iaagaalawpahgaganeavlk-mlarigk-kenipa-fiaq-v-kdknsgvklmgfghr 314
 QY 261 FAAALNGLAGPLHGLANQEVLLWIKSVVDECGENISTEQKLDYVWTKLNSGKVVPFGGLG 320
 Db 315 vykfnfdrpraktmqctchevtelgikddplldlavel-ekia-lsddyfvarklypnvdf 372
 QY 321 VLRRKTDPRYTC-QRE-FA-LKHLP--DDPPFQLVSKLYEVVPPILLELGKVKNPWPNVDA 375
 Db 373 ysgliikamgiptsmf-tvlfavarttg 399
 QY 376 HSGVLLNHYGLTEARYYTVLFGVSRSLG 403

RESULT 6

ID W20051 standard; Protein; 657 AA.
 AC W20051;
 DT 10-SEP-1997 (first entry)


```

RESULT 9
ID W56163 standard; Protein; 738 AA.
AC W56163;
DE New DNA sequence isolated from Pinctada fucata.
KW Pinctada fucata; protein; cosmetic.
OS Pinctada fucata.
PN J10080285-A.
PD 31-MAR-1998.
PF 28-MAY-1997; 138461.
PR 15-JUL-1996; JP-184459.
PA (MIKI-) MIKIMOTO SEIYAKU KK.
DR WPI: 98-254410/23.
DR N-PSDB: V22683.
PT New cDNA and e.g. vector, host cell and polypeptide - used to
PT produce polypeptide in high yields, which is used in cosmetics
PS Claim 9; Pages 9-11; 15pp; Japanese.
CC The present sequence represents protein encoded by a new DNA sequence
CC isolated from Pinctada fucata. The protein be used as an ingredient
CC in cosmetics.
SQ Sequence 738 AA;

Query Match 3.1%; Score 100; DB 31; Length 738;
Best Local Similarity 34.4%; Pred. No. 2.08e+01;
Matches 11; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

Db 10 lvgfssalsfgcnyrvplgfnsgymlgglrlf 41
QY 199 VIPLDDSLDYGNGFAHMLGFDSPQMLEMRLY 230

RESULT 10
ID R89749 standard; Protein; 219 AA.
AC R89749;
DE 03-SEP-1996 (first entry)
DE AFT-1 interacting protein (partial proteasome).
KW Arabidopsis fourteen-three-three 1; AFT1; Cruciferae; regulation;
KW transcription activator; gene expression modulator; crucifer;
KW signal transduction; defence response; gene therapy; transgenic;
KW plant.
OS Arabidopsis sp.
PN EP-693554-AI.
PD 24-JAN-1996.
PF 22-JUN-1995; 109669.
PR 23-JUN-1994; US-266451.
PA (GEO) GEN HOSPITAL CORP.
PA (MASS-) MASSACHUSETTS GEN HOSPITAL.
PI Goodman HM, Zhang H;
DR WPI: 96-078601/09.
DR N-PSDB: T11341.
PT Recombinant transcriptional activator AFT-1 - isolated from
PT Arabidopsis and useful for modifying gene expression and modulating
PT plant defence responses to pathogens
PS Disclosure; Page 25-26; 47pp; English.
CC R89749 is a partial Arabidopsis proteasome, a plant defence-related
CC protein (PDRP). This protein and other PDRPs were isolated using a
CC LexA/AFT1 33-194 fusion protein, i.e. AFT1 (Arabidopsis 14-3-3 1)
CC protein residues 33-194 fused to LexA, a DNA binding protein, in an
CC interaction trap assay. AFT1 is a transcriptional activator protein
CC and may be used to enhance, control or modify plant gene expression,
CC e.g. in the regulation of plant storage components (esp. napin,
CC legumin or phaseolin). AFT1 can also be used to modulate signal
CC transduction events involved in plant defence responses to pathogens
CC such as fungi, nematodes, bacteria and viruses.
SQ Sequence 219 AA;

Query Match 2.9%; Score 93; DB 17; Length 219;
Best Local Similarity 45.2%; Pred. No. 6.36e+01;
Matches 14; Conservative 8; Mismatches 7; Indels 2; Gaps 2;

Db 119 rltllkhhffsyq-ghvsaaalvlgqv-ditg 147
QY 20 RLKKIKKEFGSFGQLGNINVDVLMGMRGTC 50

RESULT 11
ID W20183 standard; protein; 237 AA.
AC W20183;
DE 08-JUL-1997 (first entry)
DE H. pylori cell envelope transporter protein, 179677 aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
KW cell envelope; transporter.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT misc_difference 126
FT misc_difference 127 /note= "encoded by codon TTR"
FT misc_difference 127 /note= "encoded by codon GTR"
PN W09640893-AI.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
DR WPI: 97-052306/05.
DR N-PSDB: T67415.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Pages 401; 1481pp; English.
CC The present sequence is a Helicobacter pylori cell envelope transporter
CC protein. The protein may be used in a vaccine to prevent or treat
CC H. pylori infection or to identify H. pylori life cycle activators or
CC compounds, useful as potential H. pylori life cycle activators or
CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
CC determined from overlapping contigs generated by mechanically shearing
CC the bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts.
SQ Sequence 237 AA;

Query Match 2.9%; Score 94; DB 21; Length 237;
Best Local Similarity 24.1%; Pred. No. 5.44e+01;
Matches 27; Conservative 33; Mismatches 44; Indels 8; Gaps 7;

Db 85 vaptfalaImptIndlvlgfvdsmgItPknhayIgyIapvflvIvrcqgvavagelpga 144
QY 184 VVAAYVRYMYKNGQVIFLID-DSLDDYGGFAHMLGFDSPQMLEMRLY-VTIHSDHEGG 240

Db 145 wvfvh-eHapggq-kntygiflta-svvsqllslvglygmfdkpvved 193
QY 241 NVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHG-LANQEV-LLMKSVWDE 290

RESULT 12
ID R84104 standard; Protein; 324 AA.
AC R84104;
DE 20-MAR-1996 (first entry)
DE Equine herpesvirus (EHV4) unique short (US2) gene prod..
KW Equine herpesvirus 4; EHV4; unique short; US2; recombinant; vaccine;
KW attenuated virus; diagnosis.
OS Equine herpesvirus 4.
FH Key Location/Qualifiers
FT region 123..140
FT /note= "EHV-1, EHV-4, HSV-1, PRV, HSV-2, MDV, and
FT IBR conserved US2 region"
PN W09522607-AI.
PD 24-AUG-1995.

```





```

181 MALQVSEFOKAYEKGILHSKYWEPTYESMNLIAQVPLVAAYVYRMKNGDTPKDES 240
241 LDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
241 LDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
301 LNLGAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYVWKTLSNGKVPVPGFHHGVLRL 360
301 LNLGAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYVWKTLSNGKVPVPGFHHGVLRL 360
361 TVPRYTCOREFAMKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLNYYG 420
361 TVPRYTCOREFAMKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLNYYG 420
421 LTEARYYTVLFGVSRALGICSQLIDRALGLPLERPKSVTMEWLENQCKKA 471
421 LTEARYYTVLFGVSRALGICSQLIDRALGLPLERPKSVTMEWLENQCKKA 471

RESULT 2 PRELIMINARY: PRT: 469 AA
ID O24135 AC O24135
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN CITL
OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SNN;
RA LA COGNATA U. LANDSCHUETZE V., WILLMITT2ER L., MUELLER-POEREP R.;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: X84226; E137432; -.
DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
DR PFAM: PF00285; citrate synt. 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 469 AA; 52492 MW; 6BE3C325 CRC32;

Query Match 89.4%; Score 3098; DB 10; Length 469;
Best Local Similarity 89.6%; Pred. No. 0.00e+00;
Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4;

Db 1 MVFYSVLLSKLSRAVQQTNLNSVRVQLQVQTSGLDLRSEL-QELPEQOQDR:KKLK 59
1 MVFYSVLLSKLSRAVQQTNLNSVRVQLQVQTSGLDLRSELQELPEQOQDR:KKLK 60
60 SEHGKVGOLGNITVDVVLGGMGRMTGLLWETSLDDPEGIRFRGLSIYECOKVLPAAKPG 119
61 SDM-KGSIGNITVDVVLGGMGRMTGLLWPHYLDPEGIRFRGLSIYECOKVLPAAKPG 119
120 EPLPEGLLLWLLTGKVPKSEQVDSLSQELRSRATVPDHV-YKTIDALPVTAHPMTQFATG 178
120 EPLPEGLLLWLLTGKVPKSEQVDSLSQELRSRATVPDHV-YKTIDALPVTAHPMTQFATG 179
179 VMAQVQSEFOKAYEKGILHSKYWEPTYESMNLIAQVPLVAAYVYRMKNGDTPKDD 238
180 VMAQVQSEFOKAYEKGILHSKYWEPTYESMNLIAQVPLVAAYVYRMKNGDTPKDD 239
239 SLDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFAA 297
240 SLDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFAA 299

```

```

298 ALNGLAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYAWKTLKSGKVPVPGFHHGVLRL 357
300 ALNGLAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYAWKTLKSGKVPVPGFHHGVLRL 359
358 KTDPRYTCOREFALKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLNYY 417
360 KTVPRYTCOREFAMKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLNYY 419
418 GLTEARYYTVLFGVSRALGICSQLIDRALGLPLERPKSVTMEWLENQCKKA 469
420 GLTEARYYTVLFGVSRALGICSQLIDRALGLPLERPKSVTMEWLENQCKKA 471

RESULT 3 PRELIMINARY: PRT: 472 AA.
ID O80433 AC O80433
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN DCCS.
OS DAUCUS CAROTA (CARROT).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; ARALIALES; APIACEAE; DAUCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. MS YONSUN;
RA TAKITA E., KOYAMA H., SHIRANO Y., SHIBATA D., HARA T.;
RL "CDNA encoding carrot mitochondrial citrate synthase.";
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: AB017159; D1033521; -.
DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 472 AA; 52656 MW; D46C38CC CRC32;

Query Match 83.4%; Score 2892; DB 10; Length 472;
Best Local Similarity 80.8%; Pred. No. 0.00e+00;
Matches 383; Conservative 55; Mismatches 30; Indels 6; Gaps 6;

Db 1 MVFYSVLLSKLSRAVQQTNLNSVRVQLQVQTSGLDLRSEL-QELPEQOQDR:KKLK 59
1 MVFYSVLLSKLSRAVQQTNLNSVRVQLQVQTSGLDLRSELQELPEQOQDR:KKLK 59
60 KAEGKVGOLGNITVDVVLGGMGRMTGLLWETSLDDPEGIRFRGLSIYECOKVLPAAKPG 119
60 KSDM-KGSIGNITVDVVLGGMGRMTGLLWPHYLDPEGIRFRGLSIYECOKVLPAAKPG 118
120 GEPLPEGLLLWLLTGKVPKSEQVDSLSQELRSRAVPEHV-YKTIDALPVTAHPMTQFAT 178
119 GEPLPEGLLLWLLTGKVPKSEQVDSLSQELRSRAVPEHV-YKTIDALPVTAHPMTQFAT 178
179 GVMALQVQSEFOKAYEKGILHSKYWEPTYESMNLIAQVPLVAAYVYRMKNGDTPKDD 238
179 GVMALQVQSEFOKAYEKGILHSKYWEPTYESMNLIAQVPLVAAYVYRMKNGDTPKDD 238
239 DSDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFA 297
239 DSDYGANFAHMLGFSSEMHELLMRLVYTHSDHEGNGVSAHTGHLVASALSDPYLSFA 298
298 AALNGLAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYVWKTLSNGKVPVPGFHHGVLRL 357
299 AALNGLAGPLHGLANQEVLLWIKSVVECGENISKEQLKDYVWKTLSNGKVPVPGFHHGVLRL 358
358 RNTDPRYTCOREFALKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLN 417
359 RNTDPRYTCOREFALKHLPEDPFLQVSKLYEVFLFQNLAKLPWPNVDHSGVLLN 417

```

```

Db 418 HYLTEARYTYVLFVGSRAIGSGLQVDRALGLPLERPKSVTMEWLENCKKS 471
QY 418 YLGLTEARYTYVLFVGSRAIGSGLQVDRALGLPLERPKSVTMEWLENCKKA 471

RESULT 4
ID O64869 PRELIMINARY; PPT: 474 AA.
AC O64869;
DT 01-AUG-1998 (TREMBREL. 07, CREATED)
DT 01-AUG-1998 (TREMBREL. 07, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN FA11.16.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS)
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KEPLAVAGE A.P., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: AC004521; G3128180; -.
DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
KW LYSASE: TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 474 AA; 52782 MW; 64FF2AC6 CRC32;

Query Match 81.48; Score 2820; DB 10; Length 474;
Best Local Similarity 79.18; Pred. No. 0.00e+00;
Matches 375; Conservative 55; Mismatches 38; Indels 6; Gaps 6;

Db 1 MVFFRSYSAFRLRSRVOGQSSLSNRWIMQSTDLDLKSL-OELIPEQQRRLKL 59
QY 1 MVFYSVLSLSKLSRA-VQOSVNSVRLVQVTSGLDLRSELVQELIPEQQRRLKI 59
Db 60 KSEHGKVLQGNITVDVIGMGRMTGLLWETSLLDPEGIRFGLSIPECQKVLPTAQSG 119
QY 60 KSDW-KGSIGNITVDVIGMGRMTGLLWPKHYLDPEGIRFGLSIPECQKVLPAKPG 118
Db 120 AEPLPEGLLWLLTGKVPKSEQVEALSVDLANRAAPVDYV-YNAIDALPSTAHPTQFAS 178
QY 119 GEPLPEGLLWLLTGKVPKSEQVNSVSGIAESGIISLIIMYTIDALPVTAHPMTQFAT 178
Db 179 GVMAVQVSEFQAYENGIIHKSFWPEPTEDCNLIARVPVAAAYVRRMYKNGSDIPSD 238
QY 179 GVMAVQVSEFQAYENGIIHKSFWPEPTEDSNLQAQVPLVAAAYVRRMYKNGDTIPKD 238
Db 239 KSDYGANFHMGLFDDEKVEL-MRLYITIHSDHEGNNVSAHTGHLVGSALSDPYLSFA 297
QY 239 ESDYGANFAMHGLFSSSEMHELLMRLYITIHSDHEGNNVSAHTGHLVGSALSDPYLSFA 298
Db 298 AALNGLAGLHGLANQVLLWIKSVVEECEDISKQLKEYVWKTLSNGKVPYGVGHVYL 357
QY 299 AALNGLAGLHGLANQVLLWIKSVVEECENISKQLDIYVWKTLSNGKVVPGFGHVL 358
Db 358 RNTDPRYVQREFALKHLPDPLFQVLSKLYEVVPLTELGLKKNPWNVDVHSGVLLN 417
QY 359 RKTVPVTCQREFAMKHLPEPLFQVLSKLYEVVPLTELQNLAKLK-PWPNVDVHSGVLLN 417
Db 418 HYLTEARYTYVLFVGSRAIGSGLQVDRALGLPLERPKSVTMEWLENCKKA 471
QY 418 YLGLTEARYTYVLFVGSRAIGSGLQVDRALGLPLERPKSVTMEWLENCKKA 471

```

```

RESULT 5
ID O96544 PRELIMINARY; PPT: 437 AA.
AC O96544;
DT 01-FEB-1997 (TREMBREL. 02, CREATED)
DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBREL. 09, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE) (OXALOACETATE TRANSACETASE)
DE (FRAGMENT).
GN CIT1.
OS BETA VULGARIS (SUGAR BEET).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC CARYOPHYLLIDAE; CARYOPHYLLALES; CHENOPODIACEAE; BETA.
RN [1]
RP SEQUENCE FROM N.A.
RA LA COGNATA U. LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC EMBL: X84228; E137433; -.
CC PFAM: PF00285; citrate synt; 1.
CC MENDEL: 15201; BETvu:2977; mn15201.
KW LYSASE.
FT NONTER 1 1
SQ SEQUENCE 437 AA; 48881 MW; 003688CE CRC32;

Query Match 76.18; Score 2637; DB 10; Length 437;
Best Local Similarity 81.38; Pred. No. 0.00e+00;
Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;

Db 1 SSNLDLRSEL-OELIPEQQRRLKKIKKEFGSFLQNGTINVDVIGMGRMTGLLWETSLLD 59
QY 35 SSGLDLRSELVQELIPEQQRRLKKIKSDMKG-SIGNITVDVIGMGRMTGLLWPKHYLD 93
Db 60 PEEGIRFGLSIPECQKVLPAASAGAPLPGLLWLLTGKVPKSEQVDSALDLKRAS 119
QY 94 PDEGIRFGLSIPECQKVLPAKPGGPEPLPEGLLWLLTGKVPKSEQVNSVSGIAESGI 153
Db 120 IPDHY-YKTIIDALPITAHPTQFCTGVMAVQVSEFQAYENGIIHKSFWPEPTEDCLSL 178
QY 154 ISLIIMYTIDALPVTAHPMTQFATGVMAVQVSEFQAYENGIIHKSFWPEPTEDSNL 213
Db 179 IAQVPLVAAAYVRRMYKNGQVPLDSDLDYDGNFAHMLGFSDSPOMLEL-MRLYVTHSDH 237
QY 214 IAQVPLVAAAYVRRMYKNGDTIPKDESIDYGANFAHMLGFSSSEMHELLMRLYVTHSDH 273
Db 238 EGGNVSHTGHLVGSPLSDPYLSFAALNGLAGLHGLANQVLLWIKSVVDEGENIST 297
QY 274 EGGNVSHTGHLVGSALSDPYLSFAALNGLAGLHGLANQVLLWIKSVVEECENISK 333
Db 298 EQLKDYVWKTLSNGKVPVPGFGLGVLRKTDPRYTQREFALKHLPDPLFQVLSKLYEVVP 357
QY 334 EQLKDYVWKTLSNGKVPVPGFGLGVLRKTDPRYTQREFAMKHLPEPLFQVLSKLYEVFL 393
Db 358 PILLELGVKKNPWNVDVHSGVLLNHVGLTEARYTYVLFVGSRAIGSGLQVDRALGLP 417
QY 394 LFLQNLAKLK-PWPNVDVHSGVLLNYITGLTEARYTYVLFVGSRAIGSGLQVDRALGLP 452
Db 418 LERPKSVTMEWLENCKPKR 435
QY 453 LERPKSVTMEWLENCKCK 470

RESULT 6
ID O24259 PRELIMINARY; PPT: 339 AA.
AC O24259;
DT 01-JAN-1998 (TREMBREL. 05, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE)
DE (OXALOACETATE TRANSACETASE).

```

```

GN C111.
OS POPULUS DELTOIDES X POPULUS BALSAMIFERA SUBSP. TRICHOCARPA.
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC VIOALES; SALICACEAE; POPULUS.
RN [1]
RP SEQUENCE FROM N.A.
RA LA COGNATA U.. LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -/- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
OXALOACETATE.
DR EMBL; X84227; E276838; -.
DR PFAM; PF00285; citrate_synt; 1.
KW LYASE.
SQ SEQUENCE 339 AA; 37834 MW; B2DA24A9 CRC32;

Query Match 59.7%; Score 2068; DB 10; Length 339;
Best Local Similarity 81.7%; Pred. No. 0.00e+00;
Matches 276; Conservative 35; Mismatches 24; Indels 3; Gaps 3;

Db 1 MPTKEQVGSALSKELDRALVHDYV-FKADALPVTAPHTQFATGVMALQVSEFOKAYE 59
| | | | |
QY 135 VPKEQVNSIVSGIAESGIISLIIMTTIDALPVTAPHTQFATGVMALQVSEFOKAYE 194
| | | | |

Db 60 KGIHKSYPEPTYESLSLIARPIVASIYIRRIYKDGKVIPIVNDLSVIGGFSMLGFD 119
| | | | |
QY 195 KGIHKSYPEPTYESDSMNLIAQVPLVAAVYVRRMYKNGDTIPKDESLSYGANFAHMLGFS 254
| | | | |

Db 120 SPEQEL-MRLYVTHSDHEGNGVSAHTGHLVASALSDPYLSFAALNGLAGPLHGLANO 178
| | | | |
QY 255 SSEHMLLMRLYVTHSDHEGNGVSAHTGHLVASALSDPYLSFAALNGLAGPLHGLANO 314
| | | | |

Db 179 EVLLWTKSVVEEGENITITOLKDYVWTKLNSGVVPGFGHGVLRKTVPRYTCQREFALK 238
| | | | |
QY 315 EVLLWTKSVVEEGENISKEQLKDYVWTKLNSGVVPGFGHGVLRKTVPRYTCQREFALK 374
| | | | |

Db 239 HLPDDPLFQLVSKLYEVPPVLTOLGKVKNPVNDVDAHSGVLLNYGLTEARYTYVLFQV 298
| | | | |
QY 375 HLPEDPLFQLVSKLYEVFLFLQNLAKLK-PWPVNDVDAHSGVLLNYGLTEARYTYVLFQV 433
| | | | |

Db 299 SRSGICSQLIDRALGLPLERPKSVTMELLENHCKA 336
| | | | |
QY 434 SRALGICSQLIDRALGLPLERPKSVTMELLENHCKA 471
| | | | |

RESULT 7
ID 075390 PRELIMINARY; PRT; 466 AA.
AC 075390;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA GOLDENTHAL M.J.;
RT "Cloning and sequence analysis of human citrate synthase cDNA.";
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -/- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
OXALOACETATE.
CC -/- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -/- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
OXIDATIVE METABOLISM.
CC -/- SIMILARITY: TO OTHER CITRATE SYNTHASES
DR EMBL; AF047042; G3288815.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 466 AA; 51706 MW; C0532604 CRC32;

Query Match 54.7%; Score 1897; DB 4; Length 466;

```

```

Best Local Similarity 59.3%; Pred. No. 0.00e+00;
Matches 256; Conservative 84; Mismatches 86; Indels 6; Gaps 6;

Db 27 SASSTNLK-DILADLPIKEQARIKTFQHQHGKTVVGOITVDMYGGMRGKGLVYETSVL 85
| | | | |
QY 34 TSSGLDLRSELVQELIPEQQDRLKKIKSDM-KGSIGNITVDMVLGGMGRGTMGLLWKP 92
| | | | |

Db 86 DPDEGIRFGFSIPECOKLLPKAKGGEPLPEGLFWLLVTGCIPTREQVSWLSKEWAKRA 145
| | | | |
QY 93 DPDEGIRFGLSIPECOKVLPAAKPGGEPLPEGLLWLLLTGKVPKQVNSIVSGIAESG 152
| | | | |

Db 146 ALPSHVY-TMLDNFTNLHPMSQLSAAVATLNSNESFAQAYARGISRTKYWELIYEDSD 204
| | | | |
QY 153 IISLIIMYTTIDALPVTAPHTQFATGVMALQVSEFOKAYEKGHKSYPEPTYESDMN 212
| | | | |

Db 205 LIAKPCVAAKIYRNLYWEGSGIGAGIDNSLDWNSHNTNMLGYTDHOFTEL-MRLYLTIHS 263
| | | | |
QY 213 LIAQVPLVAAVYVRRMYKNGDTI-PKDESLSYGANFAHMLGFSSEMHELMRLYVTHS 271
| | | | |

Db 264 DHEGNGVSAHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANOEVLVLTOLQKEVGKV 323
| | | | |
QY 272 DHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANOEVLLWIKSVVEEGENI 331
| | | | |

Db 324 SDEKLRYDINWTLNSGRVPGYGHAVLRKTDPRYTCQREFALKHLPNDPKFLVAQLYK 383
| | | | |
QY 332 SKEQLKDYVWTKLNSGVVPGFGHGVLRKTVPRYTCQREFAMKHLPEPLFQLVSKLYEV 391
| | | | |

Db 384 VPNVLLSQGKAKNPVNDVDAHSGVLLNYGLTEARYTYVLFQVSRALGVLAQLIWSBALG 443
| | | | |
QY 392 FLFLQNLAKLK-PWPVNDVDAHSGVLLNYGLTEARYTYVLFQVSRALGVLSQIWDPALG 450
| | | | |

Db 444 FPLERPKSMSTE 455
| | | | |
QY 451 LPLERPKSVIME 462
| | | | |

RESULT 8
ID Q53115 PRELIMINARY; PRT; 411 AA
AC Q53115;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (FRAGMENT).
GN GUTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB BACTERIUM;
RA RAOULT D., RYDKINA E., ROUX V., EREMEEVA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U59712; G1390003; -.
DR PFAM; PF00285; citrate_synt; 1.
FT NON_TER 1
FT NON_TER 411
SQ SEQUENCE 411 AA; 46189 MW; 9CCB3291 CRC32;

Query Match 7.8%; Score 270; DB 2; Length 411;
Best Local Similarity 28.6%; Pred. No. 2.16e-28;
Matches 92; Conservative 91; Mismatches 109; Indels 30; Gaps 20;

Db 88 LLIYGEPLSPSEQYNNFTKKVAVHSLVNERLHYL-FQTFCSSSHHPMAIMLAAGSL----SA 143
| | | | |
QY 129 LLLTGKVPKQVNSIVSGIAESGIISLIIMYTTIDALPVTAPHTQFATGVMALQVSE 188
| | | | |

Db 144 F---YP-DLLNFK--EADYELTAIRMAIKTPTIAAMSVK--YSIGQPFYIPDNALDFTEN 195
| | | | |
QY 189 FOKAYEKGHKSYPEPTYE-DSMNLIAQVPLVAAVYVRRMYKNGDT-IPKDESLSYGAN 246
| | | | |

Db 196 FLHMM-FATPCKYKVPNPIIKNALNKIFILHADHEQ-NASTSTVRIAGSSGANCAPACVST 253
| | | | |
QY 247 FAHMLGFS--EMHEL-LMR--LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAA 299
| | | | |

```

```

Db 254 GIASLWGPAGHGANEAVINMLKEIGSSNIPKYIAKAKNDPFLMGFGRV--YKNYD 311
QY 300 ALNGLACPLHGLANQEVLLWIKSV--VEEGGENISKEQLKDYVWKTLNSGKVPFGHGVL 357
Db 312 LRAAVLKETCKEVLKELGOLDGNPLLOIAIEL-EAIALKDEYFIERKLYPNVDFYSGIYK 370
QY 358 LRKTPRYTCOREF-AMKHLPEDFQLVSKLYEVFLLQLNLAQLKLPWPNVDAHSGVLL 416
Db 371 KAMGIP-SQMTVLFAIARTVG 391
QY 417 NYGLTEARYTYVLFVGSRALG 438

RESULT 9
ID Q59777 PRELIMINARY; PRT; 411 AA.
AC Q59777
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HA-91;
RA RAOULT D., EREMEVEA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U59731; G1390011; -
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
DR KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1 1
FT NON_TER 411 411
SQ SEQUENCE 411 AA; 45162 MW; 9921533A CRC32;

Query Match 7.6%; Score 264; DB 2; Length 411;
Best Local Similarity 28.0%; Pred. No. 2,91e-27;
Matches 90; Conservative 91; Mismatches 112; Indels 28; Gaps 20;

Db 88 LLIYGELPSGEQYNNFTKQVAHSLVNERLHYL-FQTFGSSSHPMALMAAVGSL---SA 143
QY 129 LLTGKVPKSEQVNSIVSGIAESGIISLIIMYTIDALPVTAPHTQFATGMALQVQSE 188
Db 144 F---YP-DLLNFK--EADYELTAIRMIKTIPTAAMSYK--YSISGPFYIPDNLSDFTEN 195
QY 189 FQAYEKGIHKSWEPTYE-DSMNLIAQVPLVAAYVYRMVYKNGDT-IPKDESIDYCAN 246
Db 196 FLHMFATPCTKYVNPINAKNLKIFILHADHEQ-NASTSTVRIAGSGANPACISTG 254
QY 247 FAHML-GFSSSE--MHLLMR-LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 255 IASLWGPAGHGANEAVINMLKEIGSSNIPKYIAKAKNDPFLMGFGRV--YKNYD 312
QY 301 LNLGAGPLHGLANQEVLLWIKSV-VEEC-GENISKEQLKDYVWKTLNSGKVPFGHGVL 358
Db 313 RAAVLKETCKEVLKELGOLDGNPLLOIAIEL-EAIALKDEYFIERKLYPNVDFYSGIYK 371
QY 359 RKTVPRYTCOREF-AMKHLPEDFQLVSKLYEVFLLQLNLAQLKLPWPNVDAHSGVLLN 417
Db 372 AMGIP-SQMTVLFAIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

RESULT 11
ID Q54382 PRELIMINARY; PRT; 356 AA.
AC Q54382
DT 01-JUN-1998 (TREMELREL. 06, CREATED)
DT 01-JUN-1998 (TREMELREL. 06, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PUERTO RICO;

```

```

RESULT 10
ID P77936 PRELIMINARY; PRT; 411 AA.
AC P77936
DT 01-FEB-1997 (TREMELREL. 02, CREATED)
DT 01-FEB-1997 (TREMELREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA MONTANA.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M 5\6;
RA ROUX V., RYDKINA E., EREMEVEA M., RAOULT D.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U74756; G1658283; -
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
DR KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1 1
FT NON_TER 411 411
SQ SEQUENCE 411 AA; 45266 MW; A54EA1CF CRC32;

Query Match 7.6%; Score 263; DB 2; Length 411;
Best Local Similarity 28.0%; Pred. No. 4,49e-27;
Matches 90; Conservative 91; Mismatches 112; Indels 28; Gaps 20;

Db 88 LLIYGELPSGEQYNNFTKQVAHSLVNERLHYL-FQTFGSSSHPMALMAAVGSL---SA 143
QY 129 LLTGKVPKSEQVNSIVSGIAESGIISLIIMYTIDALPVTAPHTQFATGMALQVQSE 188
Db 144 F---YP-DLLNFK--EADYELTAIRMIKTIPTAAMSYK--YSISGPFYIPDNLSDFTEN 195
QY 189 FQAYEKGIHKSWEPTYE-DSMNLIAQVPLVAAYVYRMVYKNGDT-IPKDESIDYCAN 246
Db 196 FLHMFATPCTKYVNPINAKNLKIFILHADHEQ-NASTSTVRIAGSGANPACISTG 254
QY 247 FAHML-GFSSSE--MHLLMR-LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 255 IASLWGPAGHGANEAVINMLKEIGSSNIPKYIAKAKNDPFLMGFGRV--YKNYD 312
QY 301 LNLGAGPLHGLANQEVLLWIKSV-VEEC-GENISKEQLKDYVWKTLNSGKVPFGHGVL 358
Db 313 RAAVLKETCKEVLKELGOLDGNPLLOIAIEL-EAIALKDEYFIERKLYPNVDFYSGIYK 371
QY 359 RKTVPRYTCOREF-AMKHLPEDFQLVSKLYEVFLLQLNLAQLKLPWPNVDAHSGVLLN 417
Db 372 AMGIP-SQMTVLFAIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

RESULT 11
ID Q54382 PRELIMINARY; PRT; 356 AA.
AC Q54382
DT 01-JUN-1998 (TREMELREL. 06, CREATED)
DT 01-JUN-1998 (TREMELREL. 06, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PUERTO RICO;

```



```

QY 189 FQAYEKGIHKSKEYEPTVE-DSMNLIAQVPLVAAYVYRMKNGDT-IPKDESIDYGAN 246
Db 196 FLHMFATPCTKYTNPIIKNALNKFILHADHEQ-NASTSTVRITAGSSGANPACISTG 254
QY 247 FAHML-GFSSSE--MHELLMR-LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 255 IASLWGPAGHGANEAVINMLKEIGSGSEYIPKYIAKAKDNDFPRLMGFGHRV--YKNYDP 312
QY 301 LNLGAGPLHLAQVLLWIKSV-VEEC-GENISKEQLKDYVWKTLNSGVVPGFGHVL 358
Db 313 RAAVLKETCKEVLKELGOLDNNPLLOIAIEL-EAIALKDEYFIERKLYPNVDYSGIYK 371
QY 359 RKTVPRTYTCOREF-AMKHLPEPLFQVSKLYEVFLFLQNLAKLKPWPNVDHAGSVLLN 417
Db 372 AMGIP-SOMFTVLFVIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

RESULT 14
ID Q59469
AC Q59469;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS ISRAELI TICK TYPHUS RICKETTSIA.
OC BACTERIA: PROTEOBACTERIA: ALPHA SUBDIVISION; RICKETTSTIALES;
OC RICKETTSTIAEAE; RICKETTSTIAE; RICKETTSTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IST CDC1;
RA RAOULT D., RYDKINA E., ROUX V., EREMEVA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC -!- OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC -!- OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U59727; G1389975;
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1
FT NON_TER 411
SQ SEQUENCE 411 AA; 46216 MW; 0DD40019 CRC32;

Query Match 7.5%; Score 259; DB 2; Length 411;
Best Local Similarity 28.3%; Fred. No. 2.53e-26;
Matches 91; Conservative 89; Mismatches 113; Indels 28; Gaps 20;

Db 88 LLIYGLPESGEQYNNFTKQVAHSHSLYNERLHYL-FOTFCSSSHPMALAAVGS!--SA 143
QY 129 LLLTGKVPKESQVNSVIGAESGIISLIIMYTTIDALPVTAHPMTQFATGMALQVQSE 188
Db 144 F---YP-DLLNFK--EADYELTAIRMIKPIITAAMSK--YSIGQFFIYDPSNLDFTEN 195
QY 189 FQAYEKGIHKSKEYEPTVE-DSMNLIAQVPLVAAYVYRMKNGDT-IPKDESIDYGAN 246
Db 196 FLHMFATPCTKYTNPIIKNALNKFILHADHEQ-NASTSTVRITAGSSGANPACISTG 254
QY 247 FAHML-GFSSSE--MHELLMR-LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 255 IASLWGPAGHGANEAVINMLKEIGSGSEYIPKYIAKAKDNDFPRLMGFGHRV--YKNYDP 312
QY 301 LNLGAGPLHLAQVLLWIKSV-VEEC-GENISKEQLKDYVWKTLNSGVVPGFGHVL 358
Db 313 RAAVLKETCKEVLKELGOLDNNPLLOIAIEL-EAIALKDEYFIERKLYPNVDYSGIYK 371
QY 359 RKTVPRTYTCOREF-AMKHLPEPLFQVSKLYEVFLFLQNLAKLKPWPNVDHAGSVLLN 417
Db 372 AMGIP-SOMFTVLFVIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

Search completed: Fri Oct 22 16:36:37 1999
Job time : 225 secs.

```

```

Db 372 AMGIP-SOMFTVLFVIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

RESULT 15
ID Q59779
AC Q59779;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSTIA SP.
OC BACTERIA: PROTEOBACTERIA: ALPHA SUBDIVISION; RICKETTSTIALES;
OC RICKETTSTIAEAE; RICKETTSTIAE; RICKETTSTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STRAIN S;
RA RAOULT D., RYDKINA E., ROUX V., EREMEVA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC -!- OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC -!- OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U59735; G1390015;
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1
FT NON_TER 411
SQ SEQUENCE 411 AA; 46236 MW; 50717ED2 CRC32;

Query Match 7.4%; Score 258; DB 2; Length 411;
Best Local Similarity 28.0%; Fred. No. 3.89e-26;
Matches 90; Conservative 90; Mismatches 113; Indels 28; Gaps 20;

Db 88 LLIYGLPESGEQYNNFTKQVAHSHSLYNERLHYL-FOTFCSSSHPMALAAVGS!--SA 143
QY 129 LLLTGKVPKESQVNSVIGAESGIISLIIMYTTIDALPVTAHPMTQFATGMALQVQSE 188
Db 144 F---YP-DLLNFK--EADYELTAIRMIKPIITAAMSK--YSIGQFFIYDPSNLDFTEN 195
QY 189 FQAYEKGIHKSKEYEPTVE-DSMNLIAQVPLVAAYVYRMKNGDT-IPKDESIDYGAN 246
Db 196 FLHMFATPCTKYTNPIIKNALNKFILHADHEQ-NASTSTVRITAGSSGANPACISTG 254
QY 247 FAHML-GFSSSE--MHELLMR-LY-VTI-HSDHEGNGVSAHTGHLVASALSDPYLSFAAA 300
Db 255 IASLWGPAGHGANEAVINMLKEIGSGSEYIPKYIAKAKDNDFPRLMGFGHRV--YKNYDP 312
QY 301 LNLGAGPLHLAQVLLWIKSV-VEEC-GENISKEQLKDYVWKTLNSGVVPGFGHVL 358
Db 313 RAAVLKETCKEVLKELGOLDNNPLLOIAIEL-EAIALKDEYFIERKLYPNVDYSGIYK 371
QY 359 RKTVPRTYTCOREF-AMKHLPEPLFQVSKLYEVFLFLQNLAKLKPWPNVDHAGSVLLN 417
Db 372 AMGIP-SOMFTVLFVIARTVG 391
QY 418 YGLTEARYTYVLFVGSRALG 438

Search completed: Fri Oct 22 16:36:37 1999
Job time : 225 secs.

```





Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 16:55:04 1999; MasPar time 6.83 Seconds
Tabular output not generated. 749.237 Million cell updates/sec

Title: >US-08-702-718-4
Description: (1-437) from US08702718.pep
Perfect Score: 3243
Sequence: 1 SSNDLRSLEQLPEQOER.....LERPKSVTMEWLEKFKRRA 437

Scoring table: PAM 150
Gap 11

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A.COMB 2:5B.COMB 3:PT9.COMB 4:backfiles1

Statistics: Mean 33.342; Variance 155.188; scale 0.215

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	97	3.0	624	2	US-08-756-	Sequence 9, Applicatio	2.02e+01
2	93	2.9	324	3	PCT-US93-0	Sequence 4, Applicatio	3.74e+01
3	93	2.9	324	1	US-08-198-	Sequence 4, Applicatio	3.74e+01
4	93	2.9	324	3	PCT-US95-0	Sequence 4, Applicatio	3.74e+01
5	93	2.9	324	1	US-08-323-	Sequence 4, Applicatio	3.74e+01
6	93	2.9	3031	1	US-07-689-	Sequence 4, Applicatio	3.74e+01
7	92	2.8	430	2	US-08-833-	Sequence 44, Applicati	4.36e+01
8	92	2.8	430	1	US-08-306-	Sequence 44, Applicati	4.36e+01
9	92	2.8	430	1	US-08-476-	Sequence 44, Applicati	4.36e+01
10	90	2.8	509	2	US-08-737-	Sequence 27, Applicati	5.90e+01
11	92	2.8	510	2	US-08-484-	Sequence 11, Applicati	4.36e+01
12	92	2.8	510	2	US-08-345-	Sequence 11, Applicati	4.36e+01
13	92	2.8	510	1	US-08-484-	Sequence 11, Applicati	4.36e+01
14	86	2.7	119	2	US-08-491-	Sequence 10, Applicati	1.07e+02
15	89	2.7	230	2	US-08-637-	Sequence 457, Applicat	6.85e+01
16	86	2.7	355	3	PCT-US95-0	Sequence 5, Applicatio	1.07e+02
17	86	2.7	355	1	US-08-012-	Sequence 5, Applicatio	1.07e+02
18	86	2.7	355	1	US-08-450-	Sequence 5, Applicatio	1.07e+02
19	87	2.7	500	1	US-08-117-	Sequence 68, Applicati	9.24e+01
20	86	2.7	557	1	US-08-309-	Sequence 4, Applicatio	1.07e+02
21	86	2.7	557	1	US-08-608-	Sequence 2, Applicatio	1.07e+02
22	86	2.7	557	1	US-08-309-	Sequence 2, Applicatio	1.07e+02
23	86	2.7	557	1	US-08-608-	Sequence 4, Applicatio	1.07e+02

24	86	2.7	557	2	US-08-967-	Sequence 4, Applicatio	1.07e+02
25	86	2.7	557	1	US-08-608-	Sequence 4, Applicatio	1.07e+02
26	86	2.7	557	1	US-08-967-	Sequence 2, Applicatio	1.07e+02
27	86	2.7	557	1	US-08-608-	Sequence 2, Applicatio	1.07e+02
28	86	2.7	557	1	US-08-608-	Sequence 2, Applicatio	1.07e+02
29	86	2.7	557	1	US-08-608-	Sequence 4, Applicatio	1.07e+02
30	89	2.7	668	1	US-07-891-	Sequence 6, Applicatio	6.85e+01
31	84	2.6	219	2	US-08-748-	Sequence 24, Applicati	1.44e+02
32	84	2.6	259	2	US-07-857-	Sequence 51, Applicati	1.44e+02
33	85	2.6	537	2	US-08-472-	Sequence 2, Applicatio	1.24e+02
34	85	2.6	537	2	US-08-808-	Sequence 2, Applicatio	1.24e+02
35	85	2.6	1005	3	PCT-US94-0	Sequence 3, Applicatio	1.24e+02
36	85	2.6	1005	1	US-08-089-	Sequence 3, Applicatio	1.24e+02
37	85	2.6	1005	2	US-08-717-	Sequence 3, Applicatio	1.24e+02
38	85	2.6	1005	2	US-08-266-	Sequence 3, Applicatio	1.24e+02
39	85	2.6	1005	1	US-08-478-	Sequence 3, Applicatio	1.24e+02
40	85	2.6	1454	3	PCT-US91-0	Sequence 32, Applicati	1.24e+02
41	85	2.6	1454	3	PCT-US93-0	Sequence 45, Applicati	1.24e+02
42	85	2.6	1454	3	PCT-US93-0	Sequence 44, Applicati	1.24e+02
43	85	2.6	1454	3	PCT-US93-0	Sequence 48, Applicati	1.24e+02
44	85	2.6	1454	3	PCT-US93-0	Sequence 12, Applicati	1.24e+02
45	84	2.6	1794	4	5183745-6	Patent No. 5183745.	1.44e+02

ALIGNMENTS

RESULT 1
ID US-08-756-317-9 STANDARD; PRT; 624 AA.
XX xxxxxx
DT
DE Sequence 9, Application US/08756317
CC
CC Sequence 9, Application US/08756317
CC Patent No. 5849894
CC GENERAL INFORMATION:
CC APPLICANT: Clemente, Thomas E.
CC APPLICANT: Kishore, Ganesh M.
CC APPLICANT: Mitsky, Timothy A.
CC APPLICANT: Stark, David M.
CC TITLE OF INVENTION: Improved Rhodospirillum Rubrum
CC TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Arnold, White & Durkee
CC STREET: P.O. Box 4433
CC CITY: Houston
CC STATE: TX
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/756,317
CC FILING DATE: 25-NOV-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/007,693
CC FILING DATE: 29-NOV-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Patterson, Melinda L.
CC REGISTRATION NUMBER: 33,062
CC REFERENCE/DOCKET NUMBER: MOBT:008
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (713) 787-1400
CC TELEFAX: (713) 787-1440
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:

QY	224	LELMRL-YVTIHS	235	
RESULT	3			
ID	US-08-198-094-4	STANDARD;	PRT:	324 AA.
XX	AC	xxxxxx		
XX	AC	xxxxxx		
DT	DT			
XX	XX			
DE	DE			
XX	Sequence 4, Application US/08198094			
CC	Sequence 4, Application US/08198094			
CC	Patent No. 5741696			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Cochran Ph.D., Mark D			
CC	TITLE OF INVENTION: Recombinant Equine Herpesvirus			
CC	NUMBER OF SEQUENCES: 77			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: John P. White			
CC	STREET: 30 Rockefeller Plaza			
CC	CITY: New York			
CC	STATE: New York			
CC	COUNTRY: USA			
CC	ZIP: 10112			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patent In Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/198,094			
CC	FILING DATE: February 17, 1994			
CC	CLASSIFICATION: 435			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: White, John P			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (212)977-9550			
CC	TELEFAX: (212)664-0525			
CC	TELEX: 422523			
CC	INFORMATION FOR SEQ ID NO: 4:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 324 amino acids			
CC	TYPE: amino acid			
CC	TOPOLOGY: linear			
CC	MOLECULE TYPE: protein			
CC	SEQUENCE 324 AA: 36456 MW: 559982 CN;			
QY	Query Match	2.98;	Score 93;	DB 1; Length 324;
	Best Local Similarity	24.7%;	Pred. No. 3.74e+01;	
	Matches 18;	Conservative	24;	Mismatches 25; Indels 6; Gaps
Db	156 NKSMPSEFV-CGKLFEIPITTV-DYKHLKQ-KVLPQGDHPESARSLQHKSFVSFVSPPP	212		
QY	165 SKWEPTVECLSLIAQFVVAAYVRYMYKQGVIPLDDSLDYGNGF-AHMLGFDSPQM	223		
Db	213 -NFKRLIYAVDP	224		
QY	224 LELMRL-YVTIHS	235		
RESULT	4			
ID	PC-T-US95-02087-4	STANDARD;	PRT:	324 AA.
XX	xxxxxx			
XX	AC			
DT	DT			
XX	XX			
DE	Sequence 4, Application PC/TUS9502087			
XX	Sequence 4, Application PC/TUS9502087			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Cochran, Mark D			

APPLICATION NUMBER: US/07/689,008
 FILING DATE: 19910422
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 337,194
 FILING DATE: 12-APR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 496,236
 FILING DATE: 23-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Lisabeth Feix
 REGISTRATION NUMBER: 31547
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 393-2000
 TELEFAX: (415) 393-2286
 TELEX: 340817 MACPAG SFO
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3031 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE 3031 AA: 328497 MW: 44365167

RESULT	7		
ID	US-08-833-485-44	STANDARD;	PRT; 430 AA.
XX			
AC	xxxxxx		
AD			
DT			
DD			
DE	Sequence 44, Application US/088333485		
XX			
CC	Sequence 44, Application US/088333485		
CC	Patent No. 5804425		
CC	GENERAL INFORMATION:		
CC	APPLICANT: Barry, Gerard F.		
CC	APPLICANT: Kishore, Ganesh M.		
CC	APPLICANT: Stedette, Stephen R.		
CC	APPLICANT: Stallings, William C.		
CC	TITLE OF INVENTION: Glyphosate Tolerant		
CC	TITLE OF INVENTION: 5-Enolpyruvylshikimate-3		
CC	NUMBER OF SEQUENCES: 69		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto		
CC	STREET: 700 Chesterfield Village Parkway		
CC	CITY: St. Louis		
CC	STATE: Missouri		
CC	COUNTRY: USA		
CC	ZIP: 63198		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: PatentIn Release #1.0, Version #		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/833,485		
CC	FILING DATE: 07-APR-1997		
CC	CLASSIFICATION: 435		
CC	PRIOR APPLICATION DATA:		
CC			

[illegible]

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/749,611
CC FILING DATE: 28-AUG-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/576,537
CC FILING DATE: 31-AUG-1990
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hoerner Jr., Dennis R.
CC REGISTRATION NUMBER: 30,914
CC REFERENCE/DOCKET NUMBER: 38-21(10660)A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (314)537-6099
CC TELEFAX: (314)537-6047
CC INFORMATION FOR SEQ ID NO: 44:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 430 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 430 AA; 46852 MW; 962324 CN;

Query Match 2.8%; Score 92; DB 1; Length 430;
Best Local Similarity 31.9%; Pred. No. 4.36e+01;
Matches 23; Conservative 15; Mismatches 31; Indels 3; Gaps 3;

Db 43 YKPLGDCRRTMDIFRHLGVEIKEDDEKLVTTS-PGYQVNTPHOVLYTGNSTTTLLA 101
QY 212 FAHMLGSDSPQMLELMR-LYVTHSDHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGL-A 269
Db 102 GLSLGSLGNEVYL 113
QY 270 GPLHGLANQEVYL 281

RESULT 9
ID US-08-476-008-44 STANDARD; PPT: 430 AA.
XX xxxxxx
AC xxxxxx
XX
DT
XX
DE Sequence 44, Application US/08476008
XX
XX Sequence 44, Application US/08476008
CC Patent No. 5627061
CC GENERAL INFORMATION:
CC APPLICANT: Barry, Gerard F.
CC APPLICANT: Kishore, Ganesh M.
CC APPLICANT: Padgett, Stephen R.
CC APPLICANT: Stallings, William C.
CC TITLE OF INVENTION: Glyphosate Tolerant
CC TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
CC NUMBER OF SEQUENCES: 69
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
CC STREET: 700 Chesterfield Village Parkway
CC CITY: St. Louis
CC STATE: Missouri
CC COUNTRY: USA
CC ZIP: 63198
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/476,008
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/306,063
CC FILING DATE: 13-SEP-1994
CC APPLICATION NUMBER: US 07/749,611
CC FILING DATE: 28-AUG-1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/576,537
CC FILING DATE: 31-AUG-1990
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hoerner Jr., Dennis R.
CC REGISTRATION NUMBER: 30,914
CC REFERENCE/DOCKET NUMBER: 38-21(10660)A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (314)537-6099
CC TELEFAX: (314)537-6047
CC INFORMATION FOR SEQ ID NO: 44:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 430 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 430 AA; 46852 MW; 962324 CN;

Query Match 2.8%; Score 92; DB 1; Length 430;
Best Local Similarity 31.9%; Pred. No. 4.36e+01;
Matches 23; Conservative 15; Mismatches 31; Indels 3; Gaps 3;

Db 43 YKPLGDCRRTMDIFRHLGVEIKEDDEKLVTTS-PGYQVNTPHOVLYTGNSTTTLLA 101
QY 212 FAHMLGSDSPQMLELMR-LYVTHSDHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGL-A 269
Db 102 GLSLGSLGNEVYL 113
QY 270 GPLHGLANQEVYL 281

RESULT 10
ID US-08-737-524B-27 STANDARD; PRT: 509 AA.
XX xxxxxx
AC xxxxxx
XX
DT
XX
DE Sequence 27, Application US/08737524B
XX
XX Sequence 27, Application US/08737524B
CC Patent No. 5912414
CC GENERAL INFORMATION:
CC APPLICANT: CARL SAVERIO FALCO
CC APPLICANT: DOMINICK ANTHONY GUIDA, JR.
CC APPLICANT: MARY ELIZABETH HARNETT LOCKE
CC TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS, CHIMERIC
CC TITLE OF INVENTION: GENES AND METHODS FOR INCREASING
CC TITLE OF INVENTION: THE METHIONINE CONTENT OF THE SEEDS
CC NUMBER OF SEQUENCES: 27
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E. J. DU PONT DE NEMOURS AND COMPANY
CC STREET: 1007 MARKET STREET
CC CITY: WILMINGTON
CC STATE: DELAWARE
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 19898
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: MICROSOFT WINDOWS 95
CC SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/737,524B
CC FILING DATE:
CC CLASSIFICATION: 800

CC ATTORNEY/AGENT INFORMATION:
CC NAME: LYNNE M. CHRISTENBURY
CC REGISTRATION NUMBER: 30,971
CC REFERENCE/DOCKET NUMBER: BB-1059-A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-992-5481
CC TELEFAX: 302-773-0164
CC TELEX: 835420
CC INFORMATION FOR SEQ ID NO: 27:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 509 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 509 AA: 54816 MW; 1292600 CN;
Query Match 2.8%; Score 90; DB 2; Length 509;
Best Local Similarity 25.9%; Pred. No. 5.90e+01;
Matches 15; Conservative 18; Mismatches 20; Indels 5; Gaps 4;
Db 415 IAKSQMTGF-GGVVSFEVAGDFATKRFIDSV-KIPYHAPFSGGCSIIDQ-PAIMSY 469
QY 133 ITAHPMTQFCVGMALQTRSEFQ--KAYEKGIHKSFEWPTYEDCLSLIAQVPPVAAAY 188
RESULT 11
ID US-08-484-494-11 STANDARD: PRT: 510 AA.
XX xxxxxx
DT
XX
DE Sequence 11, Application US/08484494
XX
XX Sequence 11, Application US/08484494
CC Patent No. 5798239
CC GENERAL INFORMATION:
CC APPLICANT: Wilson, Peter J
CC APPLICANT: Morris, Charles P
CC APPLICANT: Anson, Donald S
CC APPLICANT: Occhiodoro, Teresa
CC APPLICANT: Bielicki, Julie
CC APPLICANT: Clements, Peter R
CC APPLICANT: Hopwood, John J
CC TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
CC TITLE OF INVENTION: IDURONATE 2-SULFATASE
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Scully, Scott, Murphy & Presser
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,494
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 991,973
CC FILING DATE: 17-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Digiglio, Frank S
CC REGISTRATION NUMBER: 31,346
CC REFERENCE/DOCKET NUMBER: 84162
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 516-742-4343

CC TELEFAX: 516-742-4366
CC TELEX: 230 901 SANS UR
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 510 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 510 AA: 57860 MW; 1366158 CN;
Query Match 2.8%; Score 92; DB 2; Length 510;
Best Local Similarity 27.6%; Pred. No. 4.36e+01;
Matches 16; Conservative 16; Mismatches 23; Indels 3; Gaps 3;
Db 338 DIAGYDLNKTQMGSLPILRGASNLTWRSVLVYEGGEG-RNVTDTCTPSLSPGVS 394
QY 27 EFGSPOLGNINVD-M-VLGGMRGMTGLLWETSLLDPEGRIFRGFSIPECOKLLPAAS 82
RESULT 12
ID US-08-345-212-11 STANDARD: PRT: 510 AA.
XX xxxxxx
DT
XX
DE Sequence 11, Application US/08345212
XX
XX Sequence 11, Application US/08345212
CC Patent No. 5932211
CC GENERAL INFORMATION:
CC APPLICANT: Wilson, Peter J
CC APPLICANT: Morris, Charles P
CC APPLICANT: Anson, Donald S
CC APPLICANT: Occhiodoro, Teresa
CC APPLICANT: Bielicki, Julie
CC APPLICANT: Clements, Peter P
CC APPLICANT: Hopwood, John J
CC TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
CC TITLE OF INVENTION: IDURONATE 2-SULFATASE
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Scully, Scott, Murphy & Presser
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/345,212
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 991,973
CC FILING DATE: 17-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Digiglio, Frank S
CC REGISTRATION NUMBER: 31,346
CC REFERENCE/DOCKET NUMBER: 84162
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 516-742-4343
CC TELEX: 230 901 SANS UR
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 510 amino acids
CC TYPE: amino acid

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 510 AA; 57860 MW; 1366158 CN;

Query Match
Best Local Similarity 2.8%; Score 92; DB 2; Length 510;
Matches 16; Conservative 16; Mismatches 23; Indels 3; Gaps 3;

Db 338 DIAGYDLNKTOMDGMSSLLPILRGASNLTRSDVLVEYQEGG-RNVDTPTCPSLSPGVS 394
QY 27 EFGSFLGNINVD-M-VLGGMRGTMGLLWETSLDDPEEGIRFRGFSIPECKKLLPAAS 82

RESULT 13
ID US-08-484-493-11 STANDARD; PRT; 510 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 11, Application US/08484493
CC
Sequence 11, Application US/08484493
CC
Patent No. 5728381
CC
GENERAL INFORMATION:
CC
APPLICANT: Wilson, Peter J
CC
APPLICANT: Morris, Charles P
CC
APPLICANT: Anson, Donald S
CC
APPLICANT: Occhiodoro, Teresa
CC
APPLICANT: Bielicki, Julie
CC
APPLICANT: Clements, Peter R
CC
APPLICANT: Hopwood, John J
CC
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
CC
TITLE OF INVENTION: IDURONATE 2-SULFATASE
CC
NUMBER OF SEQUENCES: 15
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Scully, Scott, Murphy & Presser
CC
STREET: 400 Garden City Plaza
CC
CITY: Garden City
CC
STATE: New York
CC
COUNTRY: USA
CC
ZIP: 11530
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: PatentIn Release #1.0, Version #1.25
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/484,493
CC
FILING DATE: 07-JUN-1995
CC
CLASSIFICATION: 424
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: 991,973
CC
FILING DATE: 17-DEC-1992
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Digiglio, Frank S
CC
REGISTRATION NUMBER: 31,346
CC
REFERENCE/DOCKET NUMBER: 8416Z
CC
TELEPHONE: 516-742-4343
CC
TELEFAX: 516-742-4366
CC
TELEX: 230 901 SANS UR
CC
INFORMATION FOR SEQ ID NO: 11:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 510 amino acids
CC
TYPE: amino acid
CC
STRANDEDNESS: single
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
SQ SEQUENCE 510 AA; 57860 MW; 1366158 CN;

Query Match
Best Local Similarity 2.8%; Score 92; DB 1; Length 510;
Matches 16; Conservative 16; Mismatches 23; Indels 3; Gaps 3;

Best Local Similarity 27.6%; Pred. No. 4.36e+01;
Matches 16; Conservative 16; Mismatches 23; Indels 3; Gaps 3;

Db 338 DIAGYDLNKTOMDGMSSLLPILRGASNLTRSDVLVEYQEGG-RNVDTPTCPSLSPGVS 394
QY 27 EFGSFLGNINVD-M-VLGGMRGTMGLLWETSLDDPEEGIRFRGFSIPECKKLLPAAS 82

RESULT 14
ID US-08-491-845-10 STANDARD; PRT; 119 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 10, Application US/08491845
CC
Sequence 10, Application US/08491845
CC
Patent No. 5773247
CC
GENERAL INFORMATION:
CC
APPLICANT: MAEDA, Hiroaki
CC
APPLICANT: KIMACHI, Kazuhiko
CC
APPLICANT: EDA, Yasuyuki
CC
APPLICANT: SHIOSAKI, Kouichi
CC
APPLICANT: OSATOMI, Kiyoshi
CC
APPLICANT: TOKIYOSHI, Sachio
CC
TITLE OF INVENTION: RECOMBINANT ANTI-HIV ANTIBODY AND
CC
TITLE OF INVENTION: PROCESS FOR PREPARING THE SAME
CC
NUMBER OF SEQUENCES: 17
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Browdy and Neimark
CC
STREET: 419 Seventh Street N.W. Ste. 300
CC
CITY: Washington
CC
STATE: D.C.
CC
COUNTRY: USA
CC
ZIP: 20004
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: PatentIn Release #1.0, Version #1.30
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/491,845
CC
FILING DATE:
CC
CLASSIFICATION: 435
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: PCT/JP93/00039
CC
FILING DATE: 14-JAN-1993
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Browdy, Roger L.
CC
REGISTRATION NUMBER: 25,618
CC
REFERENCE/DOCKET NUMBER: MAEDA-5
CC
TELEPHONE: (202) 628-5197
CC
TELEFAX: (202) 737-3528
CC
INFORMATION FOR SEQ ID NO: 10:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 119 amino acids
CC
TYPE: amino acid
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
SQ SEQUENCE 119 AA; 13166 MW; 84737 CN;

Query Match
Best Local Similarity 2.7%; Score 86; DB 2; Length 119;
Matches 16; Conservative 13; Mismatches 26; Indels 1; Gaps 1;

Db 45 LKMGWKNNTGCESTHVEEFKGRVTMSLDTSTNTAYMELSSLRSEDYAVYVCAREY 100
QY 281 LLWIKSVVDECGENISTEQKDYVKNLNSGKVPDGLGLVLRKTDPR-YTCQREF 335

RESULT 15


```
Db 216 RVPVAAVYVRRMYKNCDSIPSDKSLDYGANFSHMLGFPDEKVKELMRLYITIHSDHEGG 275
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 QVPVAAVYVRRMYKNGQVPLDSDYDYGNGFAHMLGSPQMLMRLYITIHSDHEGG 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 276 NVSAHTGHLVGSALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVECEGDISKQOL 335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 NVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENISTEQ 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 336 KEYVWKLNSGKVIPIGYGHGVRLNTDPRYVCOREFALKHLDPDPLFQLVSKLYEVPVYL 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 301 KDYVWKLNSGKVPVPGGLGVRLKTDPRYTCOREFALKHLDPDFFQLVSKLYEVPVPL 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 396 TELGKVKNPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALER 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 361 LEUGKVKNPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALPLUER 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 PKSVTMDWLEAHCCK 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 421 PKSVTMEWLEKFCR 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
ENTRY S44316 #type complete
TITLE ethanolamine ammonia-lyase (EC 4.3.1.7) - potato
ORGANISM #formal_name Solanum tuberosum #common_name potato
DATE 08-Sep-1994 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S44316
REFERENCE S44316
#authors Landschuetz, V.; Willmitzer, L.; Mueller-Roeber, B.
#submission submitted to the EMBL Data Library, September 1993
#accession S44316
##molecule_type mRNA
##residues 1-471 ##label LAN
##cross-references EMBL:X75082; NID:g483509; PID:g483510
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS ammonia-lyase; carbon-nitrogen lyase
SUMMARY #length 471 #molecular-weight 52612 #checksum 4730

Query Match 81.3%; Score 2637; DB 2; Length 471;
Best Local Similarity 81.3%; Pred. No. 0.00e+00;
Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;

Db 35 SSGLDLRSRLVQELIPQOQRLKIKKDMKG-SIGNITVDMVLGMPGNTGLLWPHYLD 93
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 SSNDLRSRL-QELIPQOQRLKIKKKEFGSFGQGNINVDMLGMPGNTGLLWETSLLD 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 94 PDGIRFRGLSIPGKQVLPAAKPGGEPLPEGLLWLLLTGKVPKQVNSIVSGIAESGI 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 60 PEGIRPRGFSIPGKQVLPAAKPGGEPLPEGLLWLLLTGKVPKQVNSIVSGIAESGI 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 ISLLIYTTIDALPVTAPHTQFATGMALQVSEFQKAYEKIHKSKWEPTYEDSML 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 120 IPDHV-YKTIDALPVTAPHTQFATGMALQVSEFQKAYEKIHKSKWEPTYEDSCL 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 214 IAQVPLVAAVYVRRMYKNGDTPKDESLDYGANFAHMLGFSSEHMLRLYITIHSDH 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 179 IAQVPLVAAVYVRRMYKNGQVPLDSDYDYGNGFAHMLGFSPOMLEL-MPLVYVTHSDH 237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 EGGVNSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVECEGDISK 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 238 EGGVNSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENIST 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 334 EQLKDYVWKLNSGKVPVPGGLGVRLKTYPPYTCOREFALKHLDPDPLFQLVSKLYEVP 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 298 EQLKDYVWKLNSGKVPVPGGLGVRLKTYPPYTCOREFALKHLDPDFFQLVSKLYEVP 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 394 LFLQNLAKL-PWPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALGP 452
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 358 PILLELGKVKNPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALGP 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 453 LERPKSVMWLEAHCCK 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 418 LERPKSVMWLEKFCR 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
ENTRY YKUM #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor, mitochondrial -
Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
29-May-1998
ACCESSIONS JA0149; S07004
REFERENCE JA0149
#authors Unger, E.A.; Hand, J.M.; Cashmore, A.P.; Vasconcelos, A.C.
#journal Plant Mol. Biol. (1989) 13:411-418
#title Isolation of a cDNA encoding mitochondrial citrate synthase
from Arabidopsis thaliana.
#cross-references MUID:91370823
#accession JA0149
##molecule_type mRNA
##residues 1-472 ##label UNG
##cross-references EMBL:X17528; NID:gl1243; PID:c1188578; PID:g2652924
##note the sequence from Fig. 2 is inconsistent with that from
Fig. 1 in having 47-Gly, 125-Trp, 141-Leu, 151-Ser,
185-Asn, 187-Asn, 191-Asn, 348-Leu, an additional Gly
after 89-Trp, an additional Val after 393-Cys and two
additional Arg after 228-Arg, in lacking 114-Leu and
residues 233 to 241

GENETICS nuclear
#genome nuclear
COMPLEX homodimer
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; homodimer; mitochondrion;
oxo-acid-lyase; tricarboxylic acid cycle
FEATURE 343-375 #region acetyl-CoA binding #status predicted\
309,355,407 #active_site His, His, Asp #status predicted
SUMMARY #length 472 #molecular-weight 52941 #checksum 9626

Query Match 69.3%; Score 2249; DB 1; Length 472;
Best Local Similarity 71.7%; Pred. No. 0.00e+00;
Matches 314; Conservative 52; Mismatches 64; Indels 8; Gaps 8;

Db 36 STDLDLSQLELPEKHORLKLKSEHGKVLQGNITVDMVIGMRCMTGLLWETSLLDP 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 SSNDLRSRL-QELIPQOQRLKIKKKEFGSFGQGNINVDMLGMPGNTGLLWETSLLDP 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 96 BEVFALGDCRPLPECKALLPTAQSGSLNHVRRSPVASLNWKGTL-AKSKLKH-CRKTNRP 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 PEGIRPRGFSIPGKQVLPAAKPGGEPLPEGLLWLLLTGKVPKQVNSIVSGIAESGI 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 AAVSDYVYNAIDALPSTAHPTQFASGMALQVSEFQKAYEKIHKSKWEPTYEDCLN 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 118 ASIPDHVYKTIDALPVTAPHTQFATGMALQVSEFQKAYEKIHKSKWEPTYEDCLS 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 214 IAQVPLVAAVYVRRMYKNGDTPKDESLDYGANFAHMLGFSSEHMLRLYITIHSDH 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 178 IAQVPLVAAVYVRRMYKNGQVPLDSDYDYGNGFAHMLGFSPOMLEL-MPLVYVTHSDH 237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 EGGVNSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVECEGDISK 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 238 EGGVNSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENIST 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 334 EQLKDYVWKLNSGKVPVPGGLGVRLKTYPPYTCOREFALKHLDPDPLFQLVSKLYEVP 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 298 EQLKDYVWKLNSGKVPVPGGLGVRLKTYPPYTCOREFALKHLDPDFFQLVSKLYEVP 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 392 SCLTELES-EPPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALGP 450
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 358 PILLELGKVKNPVNDAGSVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGALGP 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 451 LERPKSVMWLEAHCCK 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 418 LERPKSVMWLEKFCR 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

QY 418 LERPKSVTMEWLEKPCR 435

RESULT 4
ENTRY YKPG #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 15-Oct-1982 #sequence_revision 30-Sep-1992 #text_change
18-Sep-1998

ACCESSIONS A29966; A01109; A61347
REFERENCE Evans, C.T.; Owens, D.D.; Sumegi, B.; Kispal, G.; Szere, P.A.
#authors Biochemistry (1988) 27:4680-4686
#journal Isolation, nucleotide sequence, and expression of a cDNA
#title encoding pig citrate synthase.
#cross-references MUID:89000665
#accession A29966
#molecule_type mRNA
#residues 1-464 #label EVA
#cross-references EMBL:M21197; NID:g164418; PID:g164419
A90457
REFERENCE Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Walsh, K.A.;
#authors Titani, K.
#journal Biochemistry (1982) 21:2028-2036
#title Complete amino acid sequence of porcine heart citrate
synthase.
#cross-references MUID:82231993
#accession A01109
#molecule_type protein
#residues 28-464 #label BLO
A61347
REFERENCE Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Wade, R.D.;
#authors Ericsson, L.H.; Neurath, H.; Walsh, K.A.; Titani, K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:5381-5385
#title Primary structure of porcine heart citrate synthase.
#accession A61347
#molecule_type protein
#residues 28-464 #label BL2
A92884
REFERENCE Remington, S.; Wiegand, G.; Huber, R.
#authors J. Mol. Biol. (1982) 158:111-152
#journal Crystallographic refinement and atomic models of two
different forms of citrate synthase at 2.7 and 1.7
angstroms resolution.
#cross-references MUID:83010291
#contents annotation: X-ray crystallography, 2.7 and 1.7 angstroms
COMMENT Citrate (si)-synthase is found in nearly all cells capable of
oxidative metabolism. It catalyzes the condensation of
oxaloacetate and acetyl-CoA to form citrate in the tricarboxylic
acid cycle.
COMMENT It is synthesized in the cytoplasm but functions in the
mitochondrion of eukaryote cells.
COMMENT This molecule is a dimer of identical chains. Each dimer binds two
molecules of acetyl-CoA and two molecules of oxaloacetate at two
active sites.
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; homodimer; methylated amino acid;
mitochondrion; oxo-acid-lyase; tricarboxylic acid cycle
FEATURE 1-27
#domain transit peptide (mitochondrion) #status
predicted #label SIG\
#product citrate (si)-synthase #status experimental
#label WAT\
#active site His, His, Asp #status predicted\
#modified site N6, N6, N6-trimethyllysine (Lys) #status
experimental
SUMMARY #length 464 #molecular-weight 51629 #checksum 3548
Query Match 64.6%; Score 2096; DB 1; Length 464;
Best Local Similarity 62.1%; Pred. No. 0.00e+00;
Matches 269; Conservative 79; Mismatches 84; Indels 1; Gaps 1;
Db 28 ASSTNLKDIADLIPQEARIKTRQOHGNTVVQGITVDMMYGGMRKGLVYTSVLDP 87

```

```

Db 274 AHTSHLVGSAIDPYPYLSFAAAGLAGPLHGLANQEVLFVFNKIVGEIGFNYTEREQLEW 333
Qy 244 AHTCHLVGSAIDPYPYLSFAAAGLAGPLHGLANQEVLFVFNKIVGEIGFNYTEREQLEW 303
Db 334 VKHKLSGGVWPGYGHAVLRKTDPRYECOREFALKHLPNDLFLKVLSTLYKITPGILLEQ 393
Qy 304 VKMTNSGKVPFGVLGVLRKTDPRYTCOREFALKHLPDDPFQVLVSKLYEVVPPILLEL 363
Db 394 GKAKNPWPNVDHSGVLLGYFQMTMSFTYVLFVGSRALG-CLSOLIWARGMGLPLRPK 452
Qy 364 GKVKNPWPNVDHSGVLLGHLYGLTEARYTYVLFVGSRSGLIC-SQLIWDRLGLPLRPK 422
Db 453 SHSTDGLIKL 462
Qy 423 SVTMEWLEKF 432

RESULT 6
ENTRY YKBY #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor, mitochondrial -
ALTERNATE_NAMES yeast (Saccharomyces cerevisiae)
ORGANISM protein N2019; protein YNR001c
#formal_name Saccharomyces cerevisiae
#accession 25-Feb-1985 #sequence_revision 10-Feb-1995 #text_change
DATE 26-Feb-1999
ACCESSIONS S35390; S45124; S48338; A01110; S63327
REFERENCE S35390
#authors Lindner, P.
#submission submitted to the EMBL Data Library, July 1993
#accession S35390
#molecule_type DNA
#residues 1-479 ##label LIN
##cross-references EMBL:Z3259; NID:q313749; PID:q313750
REFERENCE S45118
#authors Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
#submission submitted to the EMBL Data Library, January 1994
#description Twelve open reading frames revealed on the 23.6 kbp segment
flanking the centromere on the Saccharomyces cerevisiae
chromosome XIV right arm.
#accession S45124
#molecule_type DNA
#residues 1-479 ##label VER
##cross-references EMBL:X77395; NID:g496717; PID:g496718
REFERENCE S48338
#authors Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
#journal Yeast (1994) 10:1355-1361
#title Twelve open reading frames revealed in the 23.6 kb segment
flanking the centromere on the Saccharomyces cerevisiae
chromosome XIV right arm.
#cross-references MUID:95208356
#accession S48338
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-479 ##label VE2
##cross-references EMBL:X77395; NID:g496717; PID:g496718
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1994
REFERENCE A01110
#authors Suissa, M.; Suda, K.; Schatz, G.
#journal EMBO J. (1984) 3:1773-1781
#title Isolation of the nuclear yeast genes for citrate synthase and
fifteen other mitochondrial proteins by a new screening
method.
#cross-references MUID:85003587
#accession A01110
#molecule_type DNA
#residues 1-57,'Q',59-77,'E',78-479 ##label SUI
##cross-references GB:X00782; NID:g3602; PID:g3603
#experimental_source strain D273-10B
REFERENCE S62910
#authors Aert, R.; Verhasselt, P.; Voet, M.; Volckaert, G.
#submission submitted to the Protein Sequence Database, April 1996

```

```

#accession S63327
#molecule_type DNA
#residues 1-479 ##label AER
##cross-references EMBL:Z71616; NID:q1302468; PID:e233778; PID:q1302469;
MIPS:YNR001c
#experimental_source strain S288C
GENETICS
#gene SGD:CIT1
##cross-references SGD:S0005284; MIPS:YNR001c
#map_position 14R
#genome nuclear
COMPLEX homodimer
FUNCTION
#description carbon-carbon lyase; catalyzes condensation of oxaloacetate
and acetyl-CoA to form citrate; oxo-acid-lyase
#pathway tricarboxylic acid cycle
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS acetyl-CoA; carbon-carbon lyase; homodimer; mitochondrion;
OXO-acid-lyase; tricarboxylic acid cycle
FEATURE
1-39 #domain transit peptide (mitochondrion) #status
predicted #label TNP\
40-479 #product citrate (si)-synthase #status predicted #label
MAT\
312,358,413 #active_site His, His, Asp #status predicted
SUMMARY #length 479 #molecular-weight 53360 #checksum 4782
Query Match 56.0%; Score 1815; DB 1; Length 479;
Best Local Similarity 57.1%; Pred. No. 0.00e+00;
Matches 244; Conservative 84; Mismatches 99; Indels 0; Gaps 0;
Db 40 ASEOTLKREFAEIIIPAKAEIKFKKEHGKTVIGEVLEQAGVGMGKIGKLVWEGSLDP 99
Qy 1 SSNDLRSEIQELLPEQWERLKKIKKEFGSFGQGNINVDVLMGKMGMTLLWETSLDLP 60
Db 100 EGIIRFGRITPEIQRELPAKAGSTPELPALFLLLTGTEIPTDAQVKALSADLAARSEI 159
Qy 61 EGIIRFGRFSIPEQCKLLPAASAGAEPLEGLLWLLTGKVPSEKQVDALSADLRKRSI 120
Db 160 PEHVLIQLDSLPRDLHPMAQFSIAVTALESSEKFAKAYAGVSKKEYSWYTFEDSLDLC 219
Qy 121 PDHVYKTIIDALPITAHPTOFCGVMALOTRSEFOKAYEKGIHKSPFWEYEDCLSLIA 180
Db 220 KLPVIASKIYRNVEKDGKITSTDPNADYGNLAQLGYNKDFIDLMKLYLTIHSDHEGG 279
Qy 181 QVPVVAAYVRYMYKNGQVLPDLSLDYGGCFAMHMLGFDSPQMLELMRLYVTHSDHEGG 240
Db 280 NVSAHTHLVGSALSSPYLSLAAGNLGAPLHGRANQEVLEWLFKLREVKGDYKSTI 339
Qy 241 NVSAHTGLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDECGENISTEQL 300
Db 340 EKYLDWTLNAGRVVPGYGHAVLRKTDPRYTCOREFALKHFPDYELFKLVSTIYEVAPGV 399
Qy 301 KDYVWKTNSGKVPFGVLGVLRKTDPRYTCOREFALKHLPDDPFQVLVSKLYEVVPPIL 360
Db 400 TKHGKTKNPWPNVDHSGVLLGYFQMTMSFTYVLFVGSRALGVLGVARAIGVLPQLIIDRAGVAPIER 459
Qy 361 LELGKVKNPWPNVDHSGVLLGHLYGLTEARYTYVLFVGSRSGLIC-SQLIWDRLGLPLER 420
Db 460 PKSFSTE 466
Qy 421 PKSVTME 427
RESULT 7
ENTRY #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7), mitochondrial -
ORGANISM Neurospora crassa
#formal_name Neurospora crassa
#accession 20-May-1994 #sequence_revision 10-Nov-1995 #text_change
DATE 08-Sep-1997
REFERENCE S41563
#accession S41563
#reference S41563

```

[illegible]

```

QY 366 VKNPWFNVAHSGVLLNHGVLTEARYTTLVFGVSRSLGCSQLIWDRLGLPLPPEPKSVT 425
Db 446 TE 447
QY 426 ME 427

RESULT 9
ENTRY #type complete
TITLE 14-nm filament protein/citrate synthase (EC 4.1.3.-) -
ALTERNATE_NAMES Tetrahymena thermophila (SGS5)
ORGANISM 49K filament-forming protein
#formal_name Tetrahymena thermophila
DATE 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change
18-Sep-1998
ACCESSIONS JC5625
REFERENCE JC5625 JN0130
#authors Takeda, T.; Watanabe, Y.; Numata, O.
#journal Biochem. Biophys. Res. Commun. (1997) 237:205-210
#title Direct demonstration of the bifunctional property of
Tetrahymena 14-nm filament protein/citrate synthase
following expression of the gene in Escherichia coli.
#accession JC5625
#molecule_type mRNA
#residues 1-462 #label TAK
REFERENCE JN0130
#authors Numata, O.; Takemasa, T.; Takagi, I.; Hirono, M.; Hirano, H.;
Chiba, J.; Watanabe, Y.
#journal Biochem. Biophys. Res. Commun. (1991) 174:1028-1034
#title Tetrahymena 14-nm filament-forming protein has citrate
synthase activity.
#cross-references MUID:91128358
#accession JN0130
#molecule_type mRNA
#residues 1-462 #label NM
#cross-references GB:D90117; NID:g217406; PID:d1014848; PID:g1688046
#note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
COMMENT This protein is involved in oral morphogenesis preceding binary
fission, and in nuclear events during fertilization, such as
formation of gametic pronuclei and zygote formation of gametic
pronuclei. It also acts as a mitochondrial enzyme, citrate
synthase.
GENETICS
#genetic_code SGC5
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; cytoskeleton; mitochondrion;
oxo-acid-lyase
FEATURE
1-21 #domain transit peptide (mitochondrion) #status
predicted #label TNP\
22-462 #product 14-nm filament protein/citrate synthase #status
experimental #label TAK
SUMMARY #length 462 #molecular-weight 52575 #checksum 9423
Query Match 44.7%; Score 1451; DB 2; Length 462;
Best Local Similarity 48.9%; Pred. No. 2,20e-269;
Matches 215; Conservative 91; Mismatches 124; Indels 10; Gaps 9;
Db 22 SQTNLAKVTAETIPQKQALKEVKEYGDKVVGQYTVKQVIGMGRGKMSDLSCDPY 81
QY 2 SNLDLRSLOELPEQOERLKKIKKEFGSFLQGNVDMVLGMRGTMGLLWETSLDPE 61
Db 82 QGIFRGYTIPOLEKEFLPADPKAAQANQOELPEGFILWLLMTGOLPTHAQVDALKHEWO 141
QY 62 EGIFPFGFSIPECKLI.P-A-A-SAG-A--EPLPEGLWLLLTGKVPKSEQVDALSADLP 115
Db 142 NRGTVNQCNVNFIPLPKDLHSMIMLSMALLYLQDKSFAKLYDEGKISKKDYWEFFED 201
QY 116 KRASIPDHVYKTTDALPITAHMPMTQFCGVMALQTRSEPKAYEKG-THKSKFEWPTVED 174
Db 202 SMDLIKIPRAAIIYRHRYRDSKLDSDSKLDWAGNYAHMMGFQHVYKCEIRGLSH 261

```

```

QY 175 CLSLIAQPPVAAVYVRYMTKNGQVPLDSDSLDYGNGFAHMLGFDSPQMLELRVLTH 234
Db 262 CDHEGGNVSAHTTHLVGSALSDDPYLSYSAGVNGLAGPLHGLANQEVKLWLLQFIEKGTK 321
QY 235 SDHEGGNVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVDEGEN 294
Db 322 VSDKDIEDYVDHVISSGRVVPYGVGHAVLRDTDFRHHQVDFSKFHLKDDOMIKLLHQCAD 381
QY 295 ISTEQLKDYVMKLNLSGVKVPVGFGLVLRKTDPRYTCQREFALKHLPPDDPFFGLVSKLYE 354
Db 382 VIPKKLLTYKKIANPYPNVDCHSGVLLYSLGLTEYQYTVVFAVSRAIG-CMANLWISRA 440
QY 355 VVPILLELQGVKNPWPNDVHSGVLLNHGVLTEARYTTLVFGVSRSLGIC-SQLIWDRA 413
Db 441 FGLPIERPGSADLKWFDHY 460
QY 414 LGLPLERPKSVTMEWL-EKF 432

RESULT 10
ENTRY #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES protein LP21w; protein YP9723.01; protein YPR001w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 19-May-1995 #sequence_revision 01-Sep-1995 #text_change
06-Feb-1998
ACCESSIONS S52814; S57724; S59746
REFERENCE S52814
#authors Pearson, D.; Bowman, S.
#submission submitted to the EMBL Data Library, April 1995
#accession S52814
#molecule_type DNA
#residues 1-486 #label PEA
#cross-references EMBL:Z48951; NID:g762999; PID:g763000; MIPS:YPR001w
#experimental_source strain AB972
REFERENCE S57724
#authors Jia, Y. K.; Becam, A. M.; Slonimski, P. P.; Herbert, C. J
#submission submitted to the EMBL Data Library, June 1995
#accession S57724
#molecule_type DNA
#residues 1-486 #label JIA
#cross-references EMBL:X88846; NID:g895888; PID:g895889
REFERENCE S59746
#authors Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.;
Hall, J.; Storms, R. K.; Vo, D. H.; Winnett, E.
#submission submitted to the EMBL Data Library, July 1995
#description The sequence of Saccharomyces cerevisiae chromosome XVI right
arm.
#accession S59746
#molecule_type DNA
#residues 1-486 #label WAN
#cross-references EMBL:U31900; NID:g1276597; PID:g339735; MIPS:YPR001w
GENETICS
#gene SGD:CIT3
#cross-references SGD:S0006205; MIPS:YPR001w
#map_position 16R
KEYWORDS carbon-carbon lyase; oxo-acid-lyase
SUMMARY #length 486 #molecular-weight 53811 #checksum 9998
Query Match 38.9%; Score 1261; DB 2; Length 486;
Best Local Similarity 45.2%; Pred. No. 4,20e-229;
Matches 207; Conservative 102; Mismatches 123; Indels 26; Gaps 15;
Db 24 SSALTLEALENVIPKKRDVAKKLKACYGTF-VGPFTTSSVLGGMGNGSMFWJSTSLD 82
QY 1 SSNDLRSLOELPEQOERLKKIKKEFGS-FQLGNVDMVLGMRGTMGLLWETSLD 59
Db 83 PEHGKIFQGLTIEECQNLRTGTGDNFLPESMLWLLMTGGVPTFQAAASFPELAIPG 142
QY 60 PEEGIRFGFSIPECKLI.PAASAGAEP-LPEGLLWLLTGKVPKSEQVDALSADLPKPA 118

```



```

17-Jul-1998
ACCESSIONS A35157
REFERENCE Fukaya, M.; Takemura, H.; Okumura, H.; Kawamura, Y.;
#authors Horinouchi, S.; Beppu, T.
#journal J. Bacteriol. (1990) 172:2096-2104
#title Cloning of genes responsible for acetic acid resistance in
Acetobacter aceti.
#cross-references MUID:90202732
#accession A35157
##molecule-type DNA
##residues 1-436 ##label FUK
##cross-references GB:M34830; NID:g141729; PID:g141730
homohexamer
COMPLEX
FUNCTION
#description carbon-carbon lyase; oxo-acid-lyase; catalyzes synthesis of
citrate from acetyl-CoA and oxaloacetic acid
#pathway tricarboxylic acid cycle
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS allosteric regulation; carbon-carbon lyase; homohexamer;
oxo-acid-lyase; tricarboxylic acid cycle
FEATURE
313
SUMMARY #active_site His #status predicted
#length 436 #molecular-weight 48196 #checksum 9071
Query Match 9.5%; Score 309; DB 1; Length 436;
Best Local Similarity 31.1%; Pred. No. 1.45e-34;
Matches 102; Conservative 85; Mismatches 108; Indels 33; Gaps 29;
Db 91 EYIYLLGELPNKVOYDTFTNTLTHLLHEQIRNFRRDRAHPMAILGTYGALS 150
QY 90 EGLLWLLTGKVPSEKQVDALSADLRKASIPDHVYKTI-DALPIAHPTQFCTGVMAQ 149
Db 151 A-F-YPDANDIAP-AN-ROLA--A-MRLIAKIPTIAAWYK--YTOGEAFIYPRND-LN 200
QY 150 TRSEFOKAYEKGHKSKFEPTEDCLSLIAQVPVAAVYRMYKNGQV-I-PLDDSLD 207
Db 201 YAEFLSMFEARSEPKVPNVLARANNRILILHADHEQ-NASTSTVRLAGSTGANPFAC 259
QY 208 YGNGFAHML-G-FDSPQMLE-LMR-LY-VTI-HSDHEGNGVSAHTGLVGSPLSDPYLS 260
Db 260 IAAGIALWPGHGGANEAVLK-MLARIGK-KENIPA-FAIQ-V-KDKNSGVKLMGFGHR 314
QY 261 FAAALNGIAGPLHGLANQELVILKSVVDECGENISTEQLKDYVWKTLSNGKVPVFGGLG 320
Db 315 VKNFDPRAKIMQOTCHELTGIDKDDPLDLDELAVEL-EXIA-LSDDYFVQRKLYPNVDF 372
QY 321 VLKIDPRYTC-QRE-FA-LKHLP--DDPFQVLSKLYEVVPPILLELKGKVPNPNVDA 375
Db 373 YSGILKAMGIPTSMF-TVIFAVARTTG 399
QY 376 HSGVLLNHYGLTEARYTYTVLFGVSRSLG 403
RESULT 14
ENTRY YKPSCA #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) - Pseudomonas aeruginosa
ORGANISM #formal_name Pseudomonas aeruginosa
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
17-Jul-1998
ACCESSIONS A33596
REFERENCE Donald, L.J.; Molgat, G.F.; Duckworth, H.W.
#authors J. Bacteriol. (1989) 171:5542-5550
#journal Cloning, sequencing, and expression of the gene for
NADH-sensitive citrate synthase of Pseudomonas aeruginosa.
#cross-references MUID:90008792
#accession A33596
##molecule-type DNA
##residues 1-428 ##label DON
##cross-references GB:M29728
homohexamer
COMPLEX
FUNCTION
#description carbon-carbon lyase; oxo-acid-lyase; catalyzes synthesis of
citrate from acetyl-CoA and oxaloacetic acid
#pathway tricarboxylic acid cycle
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS allosteric regulation; carbon-carbon lyase; homohexamer;
oxo-acid-lyase; tricarboxylic acid cycle
FEATURE
306
SUMMARY #active_site His #status predicted
#length 428 #molecular-weight 47681 #checksum 7548
Query Match 9.4%; Score 306; DB 1; Length 428;
Best Local Similarity 29.3%; Pred. No. 5.24e-34;
Matches 95; Conservative 95; Mismatches 98; Indels 36; Gaps 28;
Db 89 LLLNGELPTAAKQEQVGTIKNHTMVHEQL-KTFENGFRDRAHPMAVMC-GVIGALS-AF 145
QY 95 LLLTGKVPSEKQVDALSADLRKASIPDHVYKTI-DALPIAHPTQFCTGVMAQTRSE 153
Db 146 YHDSLD--ITNPKHREVS-AH--RLIAKMP?TIAAWYK--YSGEPMMYPRND-LNYAEN 197
QY 154 FOKAYEKGHKSKFEPTEDCLSLIAQVPVAAVYRMYKNGQ-VI-PLDDSLDYCGN 211
Db 198 FLHMMFNTPCETKPLSPVLAKANDRIET-LHADHEQ-NASTSTVRLAGSGCANPFACIAS 255
QY 212 FAHML-GF--DS-P--QMLE-LM-KLYVTHSDHEGNGVSAHTGLVGSPLSDPYLSFAA 263
Db 256 GIAALWPGHGGANEAVL--R-MLDEIGDVSNDIKFEKAKDKNDPKFLM-GFGHRYK 310
QY 264 ALNGLAGPLHGLANQELVILKSVVDECGENISTEQLKDYVWKTLSNGKVPVFGGLVLR 323
Db 311 NFDPRKVMKQTCDEVLOELINDPOLELAMKLEIAHPDPY-FVE-PNLYPNVDFVSGI 368
QY 324 KTDPRYTC--QR-EPALKHLP-DDPFQVLSKLYEVVPPILLELKGKVPNPNVDAHSGV 379
Db 369 ILKAIIGIPTSMF-TVIFALARTVG 391
QY 380 LLNHYGLTEARYTYTVLFGVSRSLG 403
RESULT 15
ENTRY E69658 #type complete
TITLE citrate synthase III mmqD - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS E69658
REFERENCE Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borcherdt, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Bruschi, C.V.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Choi, S.K.; Codani, J.J.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Daniel, R.A.; Denizot, F.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Ehrlich, S.D.; Emmerson, P.T.; Foulger, D.; Fritz, C.; Fujita,
M.; Fajita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.F.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haeche, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

```

Rev. M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
 Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
 Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
 Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
 B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
 Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
 Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
 Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
 Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
 Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256

#title The complete genome sequence of the Gram-positive bacterium

Bacillus subtilis.

#cross-references MUID:98044033

#accession E59658

##status preliminary; nucleic acid sequence not shown;

translation not shown

##molecule_type DNA

##residues 1-372 #label KUN

##cross-references GB:299116; GB:AL009126; NID:g2634723; PID:el185682;

PID:g2634848

##experimental_source strain 168

GENETICS

#gene

CLASSIFICATION #superfamily citrate (si)-synthase

SUMMARY #length 372 #molecular-weight 42106 #checksum 1505

Query Match 9.3%; Score 301; DB 2; Length 372;

Best Local Similarity 28.3%; Pred. No. 4.44e-33;

Matches 96; Conservative 91; Mismatches 124; Indels 28; Gaps 23;

Db 50 LVHLLLEGLPESEMETLERKINSASSLPADHL-RLLELLPEDTHPMOGLRTGLSALAG 108

QY 92 LLLWLLTGKVPKSEQVDALSADLRKRASIP-DRVYKTDIDALPITAHPTQCTGVMALQT 150

Db 109 -YDRQID-DRSPANK--ERAYQ--L-L-GKMPALTAASYRIINKKEPILPLQ-TLSYSA 159

QY 151 RSEFQAYEKGIHKSFWPEPTVEDCLSLIAQVVAAYVYRMKNGQVPLDDSLDYGG 210

Db 160 NFLYMTGKLPSSLEQIFDRSLVLYSEHMPN-STFAARVIASTHSDLYGALTCAVASL 218

QY 211 NFAHMLGFDSPQMLE--LMRLYYTIHSDHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGL 268

Db 219 KGNLHGANEAVM-YL--LLEA-KTTSDFEQILQ-T-KLKRKEKM-GFGRVYMKKMDP 271

QY 269 AGPLHGLANQEVLLWTKSVVDECGENISTEQLKDYWKTLNSGKVPGEGLGV-LRKTDP 327

Db 272 RALMKE-ALQQLCDKAG-D--HRLYEMCEAGERLMEKEKGLYPNLDYYAAPVYMWLGIP 327

QY 328 RYTCOREFALKHLPDDPFFOLVSKLYEVVPPILLELGVKNPWPVNDVDAHSGVLLNHYGLT 387

Db 328 IPLYTPIFEF-ARTSGLCARHI--EQHANNRLEPRVSYM 364

QY 388 EARYYTVLFGVSRSLGICSLIWDRLGLPLERKSVTM 426

Search completed: Fri Oct 22 16:50:35 1999

Job time : 83 secs.

W O R L D

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 16:50:52 1999; Maspar time 13.81 Seconds

Tabular output not generated. 894.613 Million cell updates/sec

Title: >US-08-702-718-4

Description: (1-437) from US08702718.pep

Perfect Score: 3243

Sequence: 1 SSNDLRSELQELIPEQER.....LERPKSVTMEWLEKFKCKRRA 437

Scoring table: PAM 150

Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot37

1:swissprot

Statistics: Mean 50.720; Variance 85.362; scale 0.594

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2817	86.9	471	1	CISY_CITMA	CITRATE SYNTHASE, MITO
2	2249	69.3	472	1	CISY_ARATH	CITRATE SYNTHASE, MITO
3	2096	64.6	464	1	CISY_PIG	CITRATE SYNTHASE, MITO
4	1993	61.5	433	1	CISY_CHICK	CITRATE SYNTHASE, MITO
5	1890	58.3	468	1	CISY_CAEEL	PROBABLE CITRATE SYNTH
6	1870	57.7	482	1	CISY_SCHEP	CITRATE SYNTHASE, MITO
7	1815	56.0	479	1	CISY_YEAST	CITRATE SYNTHASE, MITO
8	1778	54.8	474	1	CISY_EMENI	CITRATE SYNTHASE, MITO
9	1777	54.8	475	1	CISY_ASPNG	CITRATE SYNTHASE, MITO
10	1767	54.5	467	1	CISY_CANTR	CITRATE SYNTHASE, MITO
11	1761	54.3	469	1	CISY_NEUCR	CITRATE SYNTHASE, MITO
12	1701	52.5	450	1	CISY_YEAST	CITRATE SYNTHASE, MITO
13	1451	44.7	462	1	CISY_TETH	CITRATE SYNTHASE, MITO
14	1261	38.9	486	1	CISY_YEAST	CITRATE SYNTHASE, MITO
15	323	10.0	373	1	CISY_BACCO	CITRATE SYNTHASE, MITO
16	324	10.0	437	1	CISY_CORGL	CITRATE SYNTHASE, MITO
17	309	9.5	436	1	CISY_ACERC	CITRATE SYNTHASE, MITO
18	301	9.3	372	1	CISW_BACSU	CITRATE SYNTHASE, MITO
19	303	9.3	428	1	CISY_PSEAE	CITRATE SYNTHASE, MITO
20	292	9.0	397	1	CISY_SYNY3	CITRATE SYNTHASE, MITO
21	288	8.9	431	1	CISY_MYCTU	CITRATE SYNTHASE, MITO
22	266	8.2	430	1	CISY_COXBU	CITRATE SYNTHASE, MITO
23	261	8.0	423	1	CISY_ACIAN	CITRATE SYNTHASE, MITO

Query Match

86.9%; Score 2817; DB 1; Length 471;

ALIGNMENTS

RESULT	1	CISY_CITMA	STANDARD;	PRT;	471 AA.
ID	AC	P49298;	1996 (REL. 33, CREATED)		
DT	DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)		
DT	DT	01-FEB-1996	(REL. 33, LAST ANNOTATION UPDATE)		
DE	DE	CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR	(EC 4.1.3.7).		
GN	GN	CIT			
OS	OS	CITRUS MAXIMA (PUMPELO)	(CITRUS GRANDIS).		
OC	OC	EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	OC	EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			
OC	OC	SAPINDALES; RUTACEAE; CITRUS.			
RN	RN	[1]			
RP	RP	SEQUENCE FROM N.A.			
RC	RC	STRAIN-CV. SIAMESE SWEET 2240;			
RA	RA	CANEL C.;			
RL	RL	THESIS (1994), UNIVERSITY OF CALIFORNIA AT RIVERSIDE, U.S.A.			
CC	CC	-1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +			
CC	CC	OXALOACETATE.			
CC	CC	-1- PATHWAY: TRICARBOXYLIC ACID CYCLE.			
CC	CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	CC	-1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.			
CC	CC	-1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE			
CC	CC	METABOLISM.			
CC	CC	-1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.			
CC	CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	CC	use by non-profit institutions as long as its content is in no way			
CC	CC	modified and this statement is not removed. Usage by and for commercial			
CC	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	CC	or send an email to license@isb-sib.ch).			
CC	CC	EMBL; U19481; G624676; -			
DR	DR	PROSITE; PS00480; CITRATE SYNTHASE; 1.			
DR	DR	PFAM; PF00285; citrate_synt; 1.			
DR	DR	HSP; P23007; SCSC.			
KW	KW	LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.			
FT	FT	TRANSIT 1 18 MITOCHONDRION (POTENTIAL).			
FT	FT	CHAIN 19 471 CITRATE SYNTHASE.			
FT	FT	ACT_SITE 307 307 BY SIMILARITY.			
FT	FT	ACT_SITE 353 353 BY SIMILARITY.			
FT	FT	ACT_SITE 408 408 BY SIMILARITY.			
SQ	SQ	SEQUENCE 471 AA; 52183 MW; A7E3512D CRC32;			

Best Local Similarity 84.4%; Pred. No. 0.00e+00;
Matches 369; Conservative 42; Mismatches 26; Indels 0; Gaps 0;

Db 35 SADDLHSQKEMIPQOERLKKVSDGKAOQIGNITIDVIGGMRGTMGLWETSLLDP 94
QY 1 SSNDLRSELQELIPEQOERLKKIKKEFGSFGQGNINVDVIGGMRGTMGLWETSLLDP 60

Db 95 DEGIRFGLSIPCOKLLPAAPDGPGLPEGLLWLLTGKVPKQVDSGLSKELDRATV 154
QY 61 EEGIRFGRSIPCOKLLPAASAGAPLPEGLLWLLTGKVPKQVDSGLSKELDRATV 120

Db 155 PDVYKKAIDALPVSAHPTQFASGVMAQVSEFQAYEKGIHKSKEWPTSEDSLNIA 214
QY 121 PDHWYKTIDALPITAHPTQFCTGVMAQVSEFQAYEKGIHKSKEWPTSEDSLNIA 180

Db 215 RVPVAAVYVYRIYKGGKIIPKDDSLDYGNSHMLGFDPKMLMLRLVYTIHSDHGG 274
QY 181 QVPVAAVYVYRRYKNGQVPLDSDYGGNFAGMLGFDSPQMLMLRLVYTIHSDHGG 240

Db 275 NVSAHTGHLVASALSDPYLSFLAALNGLAGPLHGLAQVLLWIKSVVDECCENVTTEL 334
QY 241 NVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLAQVLLWIKSVVDECCENISTEOL 300

Db 335 KQVWKTLSGKVPVGFHGVLRKTDPRYTCQREFALKHLPDPLFQVLSKLYEYVPPIL 394
QY 301 KQVWKTLSGKVPVGFHGVLRKTDPRYTCQREFALKHLPDPLFQVLSKLYEYVPPIL 360

Db 395 TLGKVKNPVNDVDAHSGVLLNHFLGAEARYYTVLFGVSRSLGICSLIWDALGLPLER 454
QY 361 LEIGKVKNPVNDVDAHSGVLLNHFLGAEARYYTVLFGVSRSLGICSLIWDALGLPLER 420

Db 455 PKSVTLMDWTEKCKKAA 471
QY 421 PKSVTMEWLEKFCRRA 437

RESULT 2
ID C15Y.PIG STANDARD; PRT; 472 AA.
RX MEDLINE; 91370823.
RA UNGER E.A., HAND J.M., CASHMORE A.R., VASCONCELOS A.C.;
RT "Isolation of a cDNA encoding mitochondrial citrate synthase from Arabidopsis thaliana";
RL PLANT MOL. BIOL. 13:411-418(1989).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O + OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17528; E1188578; -
DR PIR; JA0149; YKMUM.

DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
DR PRAM; PF00285; citrate_synt; 1.
DR HSP; P23007; SCSC.
KW LIASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
KW MULTIGENE FAMILY.
FT TRANSIT 1 ? MITOCHONDRION.
FT CHAIN ? 472 CITRATE SYNTHASE.
FT ACT_SITE 309 309 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 407 407 BY SIMILARITY.
SQ SEQUENCE 472 AA; 52941 MW; 954AFAB1 CRC32;

Query Match 69.3%; Score 2249; DB 1; Length 472;
Best Local Similarity 71.7%; Pred. No. 0.00e+00;
Matches 314; Conservative 52; Mismatches 64; Indels 8; Gaps 8;

Db 36 STDLDLSQQLPEKHDLKLLKSEHGKVLGNITVDVIGGMRGTMGLWETSLLDP 95
QY 1 SSNDLRSELQELIPEQOERLKKIKKEFGSFGQGNINVDVIGGMRGTMGLWETSLLDP 60

Db 96 EEFVAGDCLPECOKALLPTAQSGGLNHYRRSFVSLNWKGLT-AKSKLKH-CRKTNWR 153
QY 61 EEGIRFGRSIPCOK-LLPAA-SAGAEPLPEGLL-WLLTGKVPKQVDSGLSKELDRATV 117

Db 154 AAVSDYVNAIDALPSTAHPMTQFASGVMAQVSEFQAYEKGIHKSKEWPTSEDSLN 213
QY 118 ASIPDHVYKTIIDALPITAHPTQFCTGVMAQVSEFQAYEKGIHKSKEWPTSEDS 177

Db 214 LIARVPVAAVYVYRRYKNGDSIPSDKSLDYGANFSHMLGFDPKMLMLRLVYTIHSDH 273
QY 178 LIAQVPAVAAVYVYRRYKNGQVPLDSDYGGNFAGMLGFDSPQMLMLRLVYTIHSDH 237

Db 274 EGGVNSAHTGHLVGSALSDPYLSFAAALNGLAGPLHGLAQVLLWIKSVVDECCEDISK 333
QY 238 EGGVNSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLAQVLLWIKSVVDECCENIST 297

Db 334 EQLKEYVWKTLSGKVPVGFHGVLRKTDPRYTCQREFALKHLPDPLFQ-CCKLMLK-A 391
QY 298 EQLKDYVWKTLSGKVPVGFHGVLRKTDPRYTCQREFALKHLPDPLFQVLSKLYEYV 357

Db 392 SCLTELES-EPPWNVDAHSGVLLNHFLGAEARYYTVLFGVSRSLGICSLIWDPELLLA 450
QY 358 PILLELGKVKNPVNDVDAHSGVLLNHFLGAEARYYTVLFGVSRSLGICSLIWDALGLP 417

Db 451 LERPKSVTMDWLEAHCKK 468
QY 418 LERPKSVTMEWLEKFCR 435

RESULT 3
ID C15Y.PIG STANDARD; PRT; 464 AA.
RX MEDLINE; 91370823.
RA UNGER E.A., HAND J.M., CASHMORE A.R., VASCONCELOS A.C.;
RT "Isolation of a cDNA encoding mitochondrial citrate synthase from Arabidopsis thaliana";
RL PLANT MOL. BIOL. 13:411-418(1989).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O + OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17528; E1188578; -
DR PIR; JA0149; YKMUM.

```

[3]
X-RAY CRYSTALLOGRAPHY (2.7 AND 1.7 ANGSTROMS).
MEDLINE; 83010291.
RA REMINGTON S., WIEGAND G., HUBER P.;
"Crystallographic refinement and atomic models of two different forms
of citrate synthase at 2.7- and 1.7-A resolution.";
J. MOL. BIOL. 158:111-152(1982).
[4]
MUTAGENESIS.
RP MEDLINE; 91104711.
RP ALTER G.M., CASAZZA J.P., ZHI W., NEMETH P., SRERE P.A., EVANS C.T.;
"Mutation of essential catalytic residues in pig citrate synthase.";
BIOCHEMISTRY 29:7557-7563(1990).
RL -i- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OXALOACETATE.
CC
CC -i- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC
CC -i- SUBUNIT: HOMODIMER
CC
CC -i- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC
CC -i- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
CC
CC -i- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC
EMBL; M21197; G164419; -.
DR PIR; A29966; YKPG.
DR PDB; 1CTS; 16-JUL-88.
DR PDB; 2CTS; 09-OCT-88.
DR PDB; 3CTS; 09-OCT-88.
DR PDB; 4CTS; 16-JUL-88.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PRAM; PF00285; citrate_synt; 1.
DR KW LYSASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; 3D-STRUCTURE;
TRANSIT PEPTIDE; METHYLATION.
FT TRANSIT 1 27 MITOCHONDRION
FT CHAIN 28 464 CITRATE SYNTHASE.
FT MOD_RES 395 395 METHYLATION (TRI-).
FT ACT_SITE 301 301
FT ACT_SITE 347 347
FT ACT_SITE 402 402
FT HELIX 33 55
FT TURN 56 57
FT STRAND 59 64
FT HELIX 65 69
FT TURN 70 71
FT STRAND 76 79
FT STRAND 84 86
FT TURN 87 89
FT STRAND 90 92
FT STRAND 97 97
FT HELIX 98 104
FT STRAND 107 107
FT STRAND 114 114
FT TURN 116 116
FT HELIX 117 125
FT HELIX 131 141
FT TURN 142 143
FT HELIX 149 157
FT TURN 160 161
FT TURN 164 165
FT HELIX 166 175
FT TURN 176 177
FT HELIX 181 187
FT TURN 188 188
FT TURN 192 193
FT HELIX 194 221
FT TURN 223 224

```


Db 171 ELIDRCPPTLHPMAOFSLAVTALEHDSAFKAYERGKNDKHYEYEDCDMLIAKTVPI 230
 QY 126 KTDALPITAHPTQCTGVMALOTSEFOKAYEKGIHKSKEWPEYEDCLSLIAQVPV 185
 Db 231 AGRIYRLYRGVAVPIOMDKHSYFANVVLGFPANNEEFVELMRLYLTTHADHEGNVSA 290
 QY 186 AAYVRPMYKNGQVPLDSDLDYGGNFAHMLGF-DSPQWLEMLRLYVTHSHDEGNVSA 244
 Db 291 HTGHLVGSALSSPFLSMAASLNGLAGPLHGLAQNQVNLFLITMKKEIGDDELSEETIKSYL 350
 QY 245 HTGHLVGSPLSDPYSFAAALNGLAGPLHGLAQNQVNLFLITMKKEIGDDELSEETIKSYL 304
 Db 351 WKLNSGRVVPYGVHVLKPTDPRYTAQREFALEHLKPKMPQLVSPLYEIVPGVLTENG 410
 QY 305 WKLNSGVVPGGLGVLRKTDPRYTCOREFALKHLPDPPFQVLSKLYEIVPGVLTENG 364
 Db 411 KTKNPNVNDVSHSGVLLQYGLKEQSFYTVLFGVSRVLGVSASLIWDRALGLPIERPKSF 470
 QY 365 KVKNPWNVDAHSGVLLNHYGLTEARYTVLFGVSRSLGICSLIWDRLGLPLERPKSV 424
 Db 471 STEALK 477
 QY 425 TMEWLEK 431

RESULT 7

ID CDS_YEAST STANDARD: PRT: 479 AA.
 AC P00890;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT1 OR LYS6 OR GLU3 OR YNR001C OR N2019.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-D273-10B;
 PA LINDER P., PLECKTHUN A.;
 RL SUBMITTED (JUL-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D273-10B;
 RX MEDLINE: 85003587.
 RA SUISA M., SUDA K., SCHATZ G.;
 RT "Isolation of the nuclear yeast genes for citrate synthase and
 fifteen other mitochondrial proteins by a new screening method";
 RL EMBO J. 3:1773-1781(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RX MEDLINE: 95208356.
 RA VERHASSELT P., AERT R., VOET M., VOLCKAERT G.;
 RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
 the centromere on the Saccharomyces cerevisiae chromosome XIV right
 arm";
 RL YEAST 10:1355-1361(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GRF88;
 RX MEDLINE: 95028151.
 RA LALO D., STETTLER S., MARIOTTE S., GENDREAU E., THURIAUX P.;
 RT "Organization of the centromeric region of chromosome XIV in
 Saccharomyces cerevisiae";
 RL YEAST 10:523-533(1994).
 CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.

CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z23259; G313750; -
 CC DR EMBL: X00782; G3603; -
 CC DR EMBL: X77395; G496718; -
 CC DR EMBL: Z711616; E239778; -
 CC DR PIR: A01110; YKEY.
 CC DR PIR: S3590; S35390.
 CC DR SGD: L0000341; CIT1.
 CC DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
 CC DR PFAM: PF00285; citrate_synt; 1.
 CC DR HSP: P23007; 5CSC.
 CC KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
 KW MULTIGENE FAMILY.
 FT TRANSIT 1 37 MITOCHONDRION.
 FT CHAIN 38 479 CITRATE SYNTHASE.
 FT ACT_SITE 312 312 BY SIMILARITY.
 FT ACT_SITE 358 358 BY SIMILARITY.
 FT ACT_SITE 413 413 BY SIMILARITY.
 FT CONFLICT 58 58 E -> Q (IN REF. 2).
 FT CONFLICT 78 78 E -> EE (IN REF. 2).
 SQ SEQUENCE 479 AA; 53360 MW; 66896385 CRC32;
 Query Match 56.0%; Score 1815; DB 1; Length 479;
 Best Local Similarity 57.1%; Pred. No. 0.00e+00;
 Matches 244; Conservative 84; Mismatches 99; Indels 0; Gaps 0;
 Db 40 ASBOTLKEPFAETIPAKAEETKPKKKGKTVIGEVILIQAYGMRGKGLVWEGSVLD 99
 QY 1 SSNLDLSELQELPEQOERLKIKKEFGSPQLGNINVDVVLGMRGMRGMLGWLWETSLLDP 60
 Db 100 EGRIFRGRTIPEIORELPAKESGTEPLPEALFWLLLTGEIPTDAQVKALSADLAARSEI 159
 QY 61 EGRIFRGFSIPECOKLLPAASAGAELPEGLLWLLTGKVPSEQVDALSADLRKASI 120
 Db 160 PEHVITQLDLSPLKDLHPMAQFSIAVTALASESKFAKAYAQGVSKKEYSVTFEDSLDLG 219
 QY 121 PDHVKYKTIDALPITAHPTQCTGVMALOTSEFOKAYEKGIHKSKEWPEYEDCLSLIA 180
 Db 220 KLPVTASKIYRNVPKDKSKITSTDNPADYGNLAQLLGYENKDFIDLMRLYLIHSDHEGG 279
 QY 181 QPVVVAAYVYRMVYKNGQVPLDSDLDYGGNFAHMLGF-DSPQWLEMLRLYVTHSHDEGG 240
 Db 280 NVSAHTTHLVGSALSSPFLSMAASLNGLAGPLHGLAQNQVNLFLITMKKEIGDDELSEETI 339
 QY 241 NVSAHTGHLVGSPLSDPYSFAAALNGLAGPLHGLAQNQVNLFLITMKKEIGDDELSEETI 300
 Db 340 EKYLMDTLNAGRVPGYGVHVLKPTDPRYTAQREFALEHLKPKMPQLVSPLYEIVPGVLT 399
 QY 301 KDYVWKLNSGVVPGGLGVLRKTDPRYTCOREFALKHLPDPPFQVLSKLYEIVPGVLT 360
 Db 400 TKHGKTKNPWNVDAHSGVLLNHYGLTEARYTVLFGVSRVLGVSASLIWDRALGLPIER 459
 QY 361 LELGKVKNPWNVDAHSGVLLNHYGLTEARYTVLFGVSRSLGICSLIWDRLGLPLER 420
 Db 460 PKSFSTE 466
 QY 421 PKSVTME 427
 RESULT 8
 ID CDS_EMENI STANDARD: PRT: 474 AA.
 AC 000098;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CITA.
 OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
 OC EUROTIALES; TRICHOCOMACEAE; EMERICELLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FGSC 4;
 RX MEDLINE: 97306446.
 RA PARK B.W., HAN K.H., LEE C.Y., LEE C.H., MAENG P.J.;
 RT "Cloning and characterization of the cita gene encoding the
 RT mitochondrial citrate synthase of Aspergillus nidulans";
 RL MOL. CELLS 7:290-295(1997).
 CC -|- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -|- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -|- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -|- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -|- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U89675; G2138332; -.
 DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 DR HSP: P23007; 5CSC.
 KW LYSASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
 FT CHAIN 1 35 MITOCHONDRION (POTENTIAL).
 FT ACT_SITE 36 474 CITRATE SYNTHASE.
 FT ACT_SITE 310 310 BY SIMILARITY.
 FT ACT_SITE 356 356 BY SIMILARITY.
 FT ACT_SITE 411 411 BY SIMILARITY.
 SQ SEQUENCE 474 AA: 52223 MW; D97E1E91 CRC32;

 Query Match 54.8%; Score 1778; DB 1; Length 474;
 Best Local Similarity 58.6%; Pred. No. 0.00e+00;
 Matches 248; Conservative 77; Mismatches 97; Indels 1; Gaps 1;

 Db 42 LKETFADKLPGELEKVKLRKEHGNKVGELTLDQAYGGARGVKCLVWEGSVLDSEGR 101
 QY 6 LRSELOELIPEQOERLKKIKKEFGSGFOLGNVDMVGLGMRGTGLLWETSLDPEGR 65

 Db 102 FRGLTPECKLKPAPGGEPELPEGLFWLLLTGTEVPSEQQVDRDLSAEWAARSDLPKFTE 161
 QY 66 FRGSIPECKLKPAPGGEPELPEGLFWLLLTGTEVPSEQQVDRDLSAEWAARSDLPKFTE 125

 Db 162 ELDRVPSTLHPMAQSFSLAVTALEHSAFAKATAKINGKDKYNTYTFESMDLIAKLPTI 221
 QY 126 KTIDALPITAHPTQFCTGYMALQTRSEFQKAYEKGIHKSKEWPEYEDCLSLIAQPVV 185

 Db 222 AAKTYRNVKDGKVAPIOKDYSYNLANOLGFADKNKDFVLEMLRLYTIHSDHEGGNVA 281
 QY 186 AAYYRMYKNGQVPLDDSLDYGNGFAHMLGF-DSFQMLEMLRLYTIHSDHEGGNVA 244

 Db 282 HTHLVGSAISSPMLSLAAGLNGLAGPLHGLANQEVNLMITEMKVVNGDLSQSTKDYL 341
 QY 245 HTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVNLMITEMKVVNGDLSQSTKDYL 304

 Db 342 WSLNAGRVPGYCHAVLRKTDPRYTSOREFALRKLPDDPMFKLVSOVYKIPGVLTG 401
 QY 305 WKTLSNKGVPVFGGLGVRLKTDPRYTCQREFALKHLPDDPFQVLSKLVKVEVPPILLEG 364

 Db 402 KTKNPVNDVDAHSGVLLQYCYGLTERNYTYTVLFGVSRALGVPOLIIDRAFPAIERPKSF 461
 QY 365 KVKNPVNDVDAHSGVLLNHNHGLTETARYTYTVLFGVSRALGVPOLIIDRAFPAIERPKSF 424

Db 462 STE 464
 QY 425 TME 427

 RESULT 9
 ID C15Y_ASPNG STANDARD; PRT; 475 AA.
 AC PS1044;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT-3.
 OS ASPERGILLUS NIGER.
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
 OC EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WU-2223L;
 RA OSHIDA Y., MIYAKE K., KANAYAMA S., KIRIMURA K., USAMI S.;
 RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -|- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -|- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -|- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
 CC -|- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -|- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D63376; G927641; -.
 DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 DR HSP: P23007; 5CSC.
 KW LYSASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
 FT CHAIN 1 ? MITOCHONDRION (BY SIMILARITY).
 FT ACT_SITE 310 310 BY SIMILARITY.
 FT ACT_SITE 356 356 BY SIMILARITY.
 FT ACT_SITE 411 411 BY SIMILARITY.
 SQ SEQUENCE 475 AA: 52153 MW; 48D237BC CRC32;

 Query Match 54.8%; Score 1777; DB 1; Length 475;
 Best Local Similarity 58.6%; Pred. No. 0.00e+00;
 Matches 248; Conservative 74; Mismatches 100; Indels 1; Gaps 1;

 Db 42 LKETFADKLPGELEKVKLRKEHGNKVGELTLDQAYGGARGVKCLVWEGSVLDSEGR 101
 QY 6 LRSELOELIPEQOERLKKIKKEFGSGFOLGNVDMVGLGMRGTGLLWETSLDPEGR 65

 Db 102 FRGLTPECKLKPAPGGEPELPEGLFWLLLTGTEVPSEQQVDRDLSAEWAARSDLPKFTE 161
 QY 66 FRGSIPECKLKPAPGGEPELPEGLFWLLLTGTEVPSEQQVDRDLSAEWAARSDLPKFTE 125

 Db 162 ELDRVPSTLHPMAQSFSLAVTALEHSAFAKATAKINGKDKYNTYTFESMDLIAKLPTI 221
 QY 126 KTIDALPITAHPTQFCTGYMALQTRSEFQKAYEKGIHKSKEWPEYEDCLSLIAQPVV 185

 Db 222 AAKTYRNVKDGKVAPIOKDYSYNLANOLGFADKNKDFVLEMLRLYTIHSDHEGGNVA 281
 QY 186 AAYYRMYKNGQVPLDDSLDYGNGFAHMLGF-DSFQMLEMLRLYTIHSDHEGGNVA 244

 Db 282 HTHLVGSAISSPMLSLAAGLNGLAGPLHGLANQEVNLMITEMKVVNGDLSQSTKDYL 341
 QY 245 HTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEVNLMITEMKVVNGDLSQSTKDYL 304

Db 342 WSLNACQVVPYGVCHAVLRKTDPRYVSQREFALPKLPDDPMFKLVYQVYKAPGVLTG 401
 QY 305 WKLNSGVVPGFGLVRLKTDPRYTCQREFALKHLPDDPFFQLVYEVVPPILLELG 364
 Db 402 KTKNPVNDVSHSGVLLQYVGLTEANYTVLFGVSRALGVLPOLIIDRALGAPIERPKSY 461
 QY 365 KTKNPVNDVSHSGVLLNHVGLTEARYTVLFGVSRSLGICSLQIDRALGLPLRPKSV 424
 Db 462 STN 464
 QY 425 TME 427

RESULT 10
 ID C1SY_CANTR STANDARD: PRT: 467 AA.
 AC P79024;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT.

OS CANDIDA TROPICALIS (YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCAROMYCETALES;
 OC CANDIDACEAE; CANDIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PK233;

RA UEDA M., SANUKI S., KAWACHI H., SHIMIZU K., ATOMI H., TANAKA A.;
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
 CC OXALOACETATE.

CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.

CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.

CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AB001565; G1871439; -
 DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
 DR PRAM: PF00285; citrate_synt; 1.
 DR HSSP: P23007; 5CSC.

KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 FT CHAIN 1 467 CITRATE SYNTHASE.
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT ACT_SITE 347 347 BY SIMILARITY.
 FT ACT_SITE 402 402 BY SIMILARITY.

CC SEQUENCE 467 AA; 52004 MW; FEE342A9 CRC32;
 Query Match 54.5%; Score 1767; DB 1; Length 467;
 Best Local Similarity 56.4%; Pred. No. 0.00e+00;
 Matches 239; Conservative 84; Mismatches 99; Indels 2; Gaps 2;

Db 32 LKORLEILPAKAEVQKLDKDKYVIGVLEQAYGMRGKIGLVWGSGLDPIEGIR 91
 QY 6 LRSPLQELPEQOERLKKIKKFGSGFOLGNNVDMVLGMRGTMGLWETSLDPIEGIR 65

Db 92 FRGRTIDIOKELPKAPGGEPLPEALFWLLLTGEVPTAQTRALSEFAAASALPKHYE 151
 QY 66 FRGSPICOKLLPAAAGAEPLPEGLLWLLLTGKVPSEKQVDALSADLRKRASIPDHVY 125

Db 152 ELIDRSRSHLHPMAQFSIAVTALESSESQFAKAYAGVHKSEYKYTYEDSIELAKLPIT 211
 QY 126 KTIDALPITAHPTQFTCGVMAQTRSEFOKAYEKGHKSFWEPTVEDCLS:IAQVFFV 185

Db 212 AAKYRNVFHDGKLPQAQIDSKLDYCANLASLLGFGENKEFELMRLYLTTHSDHEGNVS 271
 QY 186 AAYYRRMYKNGQV-IPDDSLDYGNGFAHMLGF-DSQPMLELMRLYVTHSDHEGNVS 243
 Db 272 AHTHLVGSALSSPFLSLAAGNLGLAGPLHGRANOQVLEWLFKLREELNGDYSKEATEKY 331
 QY 244 AHTGLVGSPLSDPYLSEFAALNGLAGPLHGLANOQVLLWIKSVVDEGGENISTEQIKDY 303
 Db 332 LWDTLNAGRVGPGYGHAVLRKTDPRYTAQREFALKHMPDYELFKLVSNIVYVAFGVFDQH 391
 QY 304 VVKTLNSGVVPGFGLVRLKTDPRYTCQREFALKHLPDDPFFQLVYEVVPPILLEL 363
 Db 392 GMTKNPWNVSHSGVLLQYVGLTEESFYTVLFGVSRALGVLPOLIIDRGLGMPLEPKS 451
 QY 364 GKVPNPVNDVSHSGVLLNHVGLTEARYTVLFGVSRSLGICSLQIDRALGLPLRPK 423
 Db 452 FSTE 455
 QY 424 VTME 427

RESULT 11
 ID C1SY_NEUCR STANDARD: PRT: 469 AA.
 AC P34085;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT-1.
 OS NEUROSPORA CRASSA.
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
 OC SORDARIACEAE; SORDARIACEAE; NEUROSPORA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-744;
 RX MEDLINE: 94104594.

RA FEREJA T., CONTRERAS E.T., OUNG T., BOWMAN E.J., BOWMAN B.J.;
 RT "Characterization of the cit-1 gene from Neurospora crassa encoding
 RT the mitochondrial form of citrate synthase."
 RL MOL. GEN. GENET. 242:105-110(1994).
 CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
 CC OXALOACETATE.

CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.

CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

CC NOT SIGNIFICANTLY EXPRESSED AFTER 6-12 HRS OF GROWTH. IT IS
 CC AFTER ENTERING THE STATIONARY PHASE OF GROWTH.

CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.

CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M84187; G168775; -
 DR PIR: S41563; S41563.
 DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
 DR PRAM: PF00285; citrate_synt; 1.
 DR HSSP: P23007; 5CSC.

KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
 FT TRANSIT 1 33 MITOCHONDRION (POTENTIAL).
 FT CHAIN 34 469 CITRATE SYNTHASE.
 FT ACT_SITE 352 352 BY SIMILARITY.
 SQ SEQUENCE 469 AA; 52002 MW; E6281B97 CRC32;

Query Match 54.3%; Score 1761; DB 1; Length 469.

Best Local Similarity 56.7%; Pred. No. 0.00e+00;
Matches 245; Conservative 81; Mismatches 103; Indels 3; Gaps 3;

Db 38 LKREAEELLPEKIKALRKHGSKVVDKVTLDQVYGARGIKLVWEGSVLDPEGIR 97
Qy 6 LRSELOELPEQOERLUKKIKRFGSQLGNINVDWVILGMRGMLTLLWETSLDDPEGIR 65
Db 98 FRGKTIPCEQELLPKAPGKPEGLFWLLTGTVEPSEQQVRDLSAEWAARSVPKFIE 157
Qy 66 FRGFSIPCEQKLLPAASAGAEPLPEGLWLLTGTGVPSKEQVDALSADLRKRASIPDHVY 125
Db 158 ELIDRCPDLHPMAQLSLAVTALEHTSSPARAYAKINKKEYWYTFEDSMOLIAKLPTI 217
Qy 126 KTIDALPITAHPTQCTGYMALQTRSEFOKAYEKGHKSKEWTEYEDCLSLIAQVPPV 185
Db 218 AARIYGVNFQGVAAVOKDYSEFANOLGFGDNKDFVELLRLLYTLTHDHEGNSA 277
Qy 186 AAYYRMYKNGQVIFLDSDLYGGFAHMGF-DSQMLEMLRLVITIHSDHEGNSA 244
Db 278 HTHLVGSALSPFLSVAAGLGLAGLHGLANQEVNLNMLTKKVIIGDLSDEAITKYL 337
Qy 245 HTGLVGSPLSDPYLSFAAALNGLAGLHGLANQEVLLWIKSVVDCGENISTEQOLKDYV 304
Db 338 WDTINAGRVVGYAHAVLRKTDPRYSAQRKFAQEHLPEDPMFQLVSQVTKIAPKVLTEHG 397
Qy 305 WKTINSGKVVPFGVLGVRKTRDPRYTCQREFALKHLPDPFFQLVSKJYEVVPPILLELG 364
Db 398 KTKNPVNDVHSGVLLQHYGLTEANYTVLFGVSRAGVLPOLIIDRAVGAPIERPKSY 457
Qy 365 KVNPNVNDVHSGVLLNHYGLTEARYTVLFGVSRSLGICSQLIWDRLGLPLERPKSV 424
Db 458 STDKWIE-ICKK 468
Qy 425 TME-WLEKCKR 435

RESULT 12
ID C1S2.YEAST STANDARD; PRT: 460 AA.
AC P08679;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, PEROXISOMAL (EC 4.1.3.7).
GN C1T2 OR YCR005C OR YCR5C OR YCR043.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCCHAROMYCETACEAE;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92254505.
RA BITEAU N., FREMAUX C., HEBRARD S., MENARA A., AIGLE M., CROUZET M.;
RT "The complete sequence of a 10.8kb fragment to the right of the
RT chromosome III centromere of *Saccharomyces cerevisiae*.";
RL YEAST 8:61-70(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87089811.
RA ROSENKRANTZ M., ALAM T., KIM K.-S., CLARK B.J., SRERE P.A.,
RA GUARENTE L.P.;
RT "Mitochondrial and nonmitochondrial citrate synthases in
RT *Saccharomyces cerevisiae* are encoded by distinct homologous genes.";
RL MOL. CELL. BIOL. 6:4509-4515(1986).
RN [3]
RP SEQUENCE OF 1-24 FROM N.A.
RX MEDLINE; 91094853.
RA LIAO X., SMALL W.C., SRERE P.A., BUTOW R.A.;
RT "Intramitochondrial functions regulate nonmitochondrial citrate
RT synthase (CIT2) expression in *Saccharomyces cerevisiae*.";
RL MOL. CELL. BIOL. 11:38-46(1991).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.

-1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
-1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; Z11113; G3299; -
DR EMBL; M14686; G171227; -
DR EMBL; M54982; G171229; -
DR EMBL; X59720; E264468; -
DR PIR; A25393; YKBYC.
DR YEPD; 5440; -
DR SGD; L0000342; CIT2.
DR PROSITE; PS00342; MICROBODIES_CTER; 1.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSSP; P23007; 5CSC.
DR LYASE; TRICARBOXYLIC ACID CYCLE; PEROXISOME; MULTIGENE FAMILY.
FT ACT_SITE 293 293 BY SIMILARITY.
FT ACT_SITE 339 339 BY SIMILARITY.
FT ACT_SITE 394 394 BY SIMILARITY.
FT SITE 458 460 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 460 AA; 51413 MW; 9E9BE2FD CRC32;
Query Match 52.5%; Score 1701; DB 1; Length 460;
Best Local Similarity 53.8%; Pred. No. 0.00e+00;
Matches 227; Conservative 90; Mismatches 105; Indels 0; Gaps 0;
Db 26 LKERSEIYPIHAODVRQFYKHEGCKTKISDVLLQVYGGKRGIPGSGWESVLDPEGIR 85
Qy 6 LRSELOELPEQOERLUKKIKRFGSQLGNINVDWVILGMRGMLTLLWETSLDDPEGIR 65
Db 86 FRGRTIADIQKDLPKAKGSSQPLPEALFWLLTGTVEPTQAQVENLSADLMSRSELPSPRV 145
Qy 66 FRGFSIPCEQKLLPAASAGAEPLPEGLWLLTGTGVPSKEQVDALSADLRKRASIPDHVY 125
Db 145 QLLDNLPKDLHPMAQFSTAVTALESSEKPAKAYAQGISKQDYWYTFEDSLDLGLKLPVI 205
Qy 126 KTIDALPITAHPTQCTGYMALQTRSEFOKAYEKGHKSKEWTEYEDCLSLIAQVPPV 185
Db 206 AAKYRNVFKDKMGEVDPNADYAKNLVNLIGSKDEDFVDMRLYTLTHSDHEGNSAH 265
Qy 186 AAYYRMYKNGQVIFLDSDLYGGFAHMGFDSQMLEMLRLVITIHSDHEGNSAH 245
Db 266 TSHLVGSALSPFLSVAAGLGLAGLHGLANQEVNLNMLTKKVIIGDLSDEAITKYLW 325
Qy 246 TGLVGSPLSDPYLSFAAALNGLAGLHGLANQEVLLWIKSVVDCGENISTEQOLKDYV 305
Db 326 DTLASGRVPIPGYHAYLRKTDPRYMAQRKFAQEHLPEDPMFQLVSQVTKIAPKVLTEHG 385
Qy 306 KTLASGVKVPFGVLGVRKTRDPRYTCQREFALKHLPDPFFQLVSKJYEVVPPILLELG 365
Db 386 TKNPNVNDVHSGVLLQHYGLTEANYTVLFGVSRAGVLPOLIIDRAVGAPIERPKSY 445
Qy 366 KVNPNVNDVHSGVLLNHYGLTEARYTVLFGVSRSLGICSQLIWDRLGLPLERPKSV 425
Db 446 TE 447
Qy 426 ME 427
RESULT 13
ID C1S2.YEAST STANDARD; PRT: 462 AA.
AC P24118;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7) (114 NM
DE FILAMENT-FORMING PROTEIN).
OS TETRAHYMENA THERMOPHILA.
OC EUKARYOTA; ALVOPOLATA; CILIOPHORA; OLIGOHYMNENOPHOREA; HYMENOSTOMATIDA;
OC TETRAHYMENA; TETRAHYMENA.
RN [1]
PP SEQUENCE FROM N.A., AND SEQUENCE OF 22-40.
RX MEDLINE: 91128358.
RA NUMATA O., TAKEKASA T., TAKAGI I., HIRONO M., HIRANO H., CHIBA J.,
RA WATANABE Y.;
RI "Tetrahymena 14-nm filament-forming protein has citrate synthase
RT activity.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 174:1028-1034(1991).
CC -1- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN ORAL MORPHOGENESIS AND
CC IN PRONUCLEAR BEHAVIOR DURING CONJUGATION. RESPIRATORY ENZYME.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX AND CYTOPLASMIC
CC CYTOSKELETON.
CC CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90117; E33571; -
CC PIR: JN0130; JN0130.
CC DR PROSITE: PS00480; CITRATE_SYNTHASE: 1.
CC DR PFAM: PF00285; citrate_synt; 1.
CC DR HSSP: P23007; 5CSC.
CC KW LYASE: TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
CC KW CYTOSKELETON.
CC TRANSIT 1 21 MITOCHONDRION.
FT CHAIN 22 462 CITRATE SYNTHASE.
FT ACT-SITE 300 300 BY SIMILARITY.
FT ACT-SITE 346 346 BY SIMILARITY.
FT ACT-SITE 401 401 BY SIMILARITY.
SQ SEQUENCE 462 AA; 52575 MW; 7B6A39BC CRC32;

Query Match 44.7%; Score 1451; DB 1; Length 462;
Best Local Similarity 48.9%; Pred. No. 0.00e+00;
Matches 215; Conservative 91; Mismatches 124; Indels 10; Gaps 9;

Db 22 SOTNLKKVIAEIIPOKQAEIKVEKRYGDKVGVGYTVKOVIGMGMKGLMSDLSRCDPY 81
QY 2 SNIDLSELOELPEQOERLKKIKKEFGSFGQIGNVDMVGLGMGNTGLWETSLDPE 61

Db 82 QGIIFRGTYPOLKEFLPADPKAADOANOEPLEGIFWLLMTGOLPHTHAQVDALKHEWQ 141
QY 62 EGRFRGFSIPECKQLP-A-A-SAG-A--EPLPEGLMLLTGKVPKSEQVDALSADLR 115

Db 142 NRGTVQDCVNFNLNPKDLHSTMTLSMALLYLQKDSKFAKLYDEGKISKDYWPFFED 201
QY 116 KRASIPDHVYKTDALPITAHPTQCTGVMALQTSFEQKAYEG- IHKSEKWEPTVED 174

Db 202 SMDLIARPVAAIIYRHKRYDSKLDSDSKLDWAGNYAHMGEFQHVYKCEIRGYLSH 261
QY 175 CLSLIAQVPVAAVYVRYMKNOQVPLDSDLDYGNFAHMLGFDSPQMLEMLRVYTH 234

Db 262 CDHEGNNVSAHTHLVGSALSDPYLSYAGVGNLAGLHGLANQEVKWLLOFIEKGTG 321
QY 235 SDHEGNNVSAHTHLVGSPLSDPYLSFAAALNLAGLHGLANQEVLLIKSWDECGEN 294

Db 322 VSKDIEDYVDHVISSRGVVPVGYGHVLRDTPDRFHHDVDESKFHLKDDQMKLLHQCAD 381

QY 295 ISTEQLDXYVWKTLNSGKVVPFGVLGRKTDPRYTQCOREFALKHLPDDFFQVLVSKLYE 354
Db 392 VIPKKLLTYKKIANDPYNVDCHSGVLLYSGLTEYQYTVVFAVSRAIG-CMANLINSRA 440
QY 355 VPPILLLELGGVKNPWPNDVHSGVLLNHVGLTEARYTVTLFGVRSRLGIC-SOLIWDR 413
Db 441 FGLPIERPGRSADLAKWFHDYK 460
QY 414 LGLPLERPKSVTMEWL-EKF 432

RESULT 14
ID CISK_YEAST STANDARD; PRT; 486 AA.
AC P43635;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE 3 (EC 4.1.3.7).
GN CIT3 OR YP001W OR YP9723.01.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CW04;
RA JIA Y.K., BECAM A.-M., SLONIMSKI P.P., HERBERT C.J.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA PEARSON D., BOWMAN S., BAPRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WANG Y., AHMED A., BUSSEY H., FORTIN N., FRIESEN J.D., HALL J.,
RA STORMS R.K., VO D.H., WINNETT E.;
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X88846; S895889; -
CC EMBL: Z48951; G763000; -
CC EMBL: Z71255; E236811; -
CC EMBL: U31900; G939735; -
CC SGD: L0002855; CIT3.
CC PROSITE: PS00342; MICROBODIES_CTER: 1.
CC PROSITE: PS00480; CITRATE_SYNTHASE; 1.
CC PFAM: PF00285; citrate_synt; 1.
CC DR HSSP: P23007; 5CSC.
CC KW LYASE: TRICARBOXYLIC ACID CYCLE; MULTIGENE FAMILY.
FT ACT-SITE 315 315 BY SIMILARITY.
FT ACT-SITE 361 361 BY SIMILARITY.
FT ACT-SITE 419 419 BY SIMILARITY.
FT SITE 484 486 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 486 AA; 53811 MW; 806F987D CRC32;

Query Match 38.9%; Score 1261; DB 1; Length 486;
Best Local Similarity 45.2%; Pred. No. 1.32e-268;
Matches 207; Conservative 102; Mismatches 123; Indels 26; Gaps 15;

Db 24 SSAITLKEALENVTPKKRDVKKLKYGVSTF-VSPITISSVLGSMKSNJSMFW-GTSLD 82

Search completed: Fri Oct 22 16:51:58 1999
Job time : 66 secs.

W P E R E H

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 16:52:16 1999; MasPar time 27.64 seconds
Tabular output not generated. 863.073 Million cell updates/sec

Title: >US-08-702-718-4
Description: (1-437) from US08702718.pep
Perfect Score: 3243
Sequence: 1 SSNLDLRSELQELPEQOER.....LERPKSVTMEWLEKFKCKRA 437

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 48.796; Variance 86.952; scale 0.561

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	3243	100.0	437	10	CITRATE (SI)-SYNTHASE	0.00e+00
2	2867	88.4	472	10	CITRATE SYNTHASE (EC 4	0.00e+00
3	2834	87.4	474	10	CITRATE SYNTHASE (EC 4	0.00e+00
4	2800	85.3	469	10	CITRATE SYNTHASE (EC 4	0.00e+00
5	2637	81.3	471	10	ETHANOLAMINE AMMONIA-L	0.00e+00
6	2166	66.8	339	10	CITRATE (SI)-SYNTHASE	0.00e+00
7	2072	63.9	466	4	CITRATE SYNTHASE (EC 4	0.00e+00
8	270	8.3	372	1	CITRATE SYNTHASE (CITZ	8.07e-31
9	256	7.9	379	1	CITRATE SYNTHASE (EC 4	5.18e-28
10	239	7.4	427	2	CITRATE SYNTHASE (EC 4	1.19e-24
11	240	7.4	431	2	CITRATE SYNTHASE (EC 4	7.59e-25
12	228	7.0	379	2	CITRATE SYNTHASE	1.67e-22
13	226	7.0	411	2	CITRATE SYNTHASE (FRAG	4.08e-22
14	223	6.9	356	2	CITRATE SYNTHASE (EC 4	1.55e-21
15	223	6.9	411	2	CITRATE SYNTHASE (EC 4	1.55e-21
16	222	6.8	411	2	CITRATE SYNTHASE (EC 4	2.41e-21
17	221	6.8	411	2	CITRATE SYNTHASE (EC 4	3.76e-21
18	219	6.8	411	2	CITRATE SYNTHASE (EC 4	9.10e-21
19	219	6.8	411	2	CITRATE SYNTHASE (EC 4	9.10e-21
20	218	6.7	411	2	CITRATE SYNTHASE (EC 4	1.42e-20

21	212	6.5	429	2	033915	CITRATE SYNTHASE (EC 4	1.98e-19
22	207	6.4	410	2	059769	CITRATE SYNTHASE (EC 4	1.76e-18
23	200	6.2	434	2	P94325	CITRATE SYNTHASE (EC 4	3.64e-17
24	179	5.5	390	2	070008	CITRATE SYNTHASE (EC 4	2.65e-13
25	161	5.0	322	2	Q59236	CITRATE SYNTHASE (EC 4	4.10e-10
26	158	4.9	315	2	Q59235	CITRATE SYNTHASE (EC 4	1.36e-09
27	158	4.9	318	2	Q59238	CITRATE SYNTHASE (EC 4	1.36e-09
28	158	4.9	319	2	Q59234	CITRATE SYNTHASE (EC 4	1.36e-09
29	156	4.8	349	2	Q30966	CITRATE SYNTHASE (EC 4	2.99e-09
30	149	4.6	321	2	Q59237	CITRATE SYNTHASE (EC 4	4.62e-08
31	145	4.5	319	2	P77822	CITRATE SYNTHASE (EC 4	2.15e-07
32	144	4.4	393	2	008395	CITRATE SYNTHASE (EC 4	3.15e-07
33	108	3.3	295	2	066147	HYPOTHETICAL 31.0 KD P	1.12e-01
34	106	3.3	689	3	074934	ACYL-COA OXIDASE 1 (EC	2.12e-01
35	103	3.2	294	2	030527	PCRV	5.47e-01
36	104	3.2	299	2	045540	CITRATE SYNTHASE (FRAG	4.00e-01
37	101	3.1	375	2	Q55851	HYPOTHETICAL 40.8 KD P	1.02e-00
38	100	3.1	378	5	P91198	CODED FOR BY C. ELEGAN	1.38e+00
39	101	3.1	522	5	Q23288	CODED FOR BY C. ELEGAN	1.02e+00
40	102	3.1	572	2	083970	DNA POLYMERASE III, SU	7.47e-01
41	99	3.1	699	3	074938	ACYL-COA OXIDASE 5 (EC	1.87e+00
42	100	3.1	738	5	002402	INSOLUBLE PROTEIN.	1.38e+00
43	102	3.1	791	14	Q83424	REPLICASE.	7.47e-01
44	99	3.1	1229	11	035836	NRD2 CONVERTASE (EC 3.	1.87e+00
45	102	3.1	1382	3	013559	YLR466WP.	7.47e-01

ALIGNMENTS

RESULT 1
ID Q96544 PRELIMINARY; PRT: 437 AA.
AC Q96544;
DI 01-FEB-1997 (TREMUREL. 02, CREATED)
DT 01-FEB-1997 (TREMUREL. 02, LAST SEQUENCE UPDATE)
DI 01-JAN-1999 (TREMUREL. 09, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE) (OXALOACETATE TRANSACETASE)
DN (FRAGMENT).
GN CIT1.
OS BETA VULGARIS (SUGAR BEET).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
CC CARYOPHYLLIDAE; CARYOPHYLLALES; CHENOPODIACEAE; BETA.
RN [1]
RP SEQUENCE FROM N.A.
PA LA COGNATA U.; LANSCHUETZE V.; WILLMITZER L.; MUELLEP-POEREP B;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
DR EMBL; X84228; E137433; -.
DR PFAM; PF00285; citrate_synt; 1.
DR MENDEL; 15201; BETvu;2977;mn15201.
KW LVASE.
FT NON_TER 1 1
SQ SEQUENCE 437 AA; 48881 MW; 003688CE CRC32;
Query Match 100.0%; Score 3243; DB 10. Length 437.
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	SSNLDLRSELQELPEQOERLKKIKKEGFSQFQGNINVDVWLGMRGNTGLLWETSLDLP	60
Qy	1	SSNLDLRSELQELPEQOERLKKIKKEGFSQFQGNINVDVWLGMRGNTGLLWETSLDLP	60
Dr	61	PEG-PFRGFSIPEQOKLLPAASAGAEPLPEGLLLWLLITGKVPSEQVDALSADLRKRASI	120
Qy	61	EGINFRGFSIPEQOKLLPAASAGAEPLPEGLLLWLLITGKVPSEQVDALSADLRKRASI	120
Db	121	PDHVKYTTIDALPITAHMPTQCTGVMALQTRSEFQKAYEKGIHKSKEWPEYEDCLSLIA	180
Qy	121	PDHVKYTTIDALPITAHMPTQCTGVMALQTRSEFQKAYEKGIHKSKEWPEYEDCLSLIA	180
Db	181	QVPVAAVYRMVKNQGVIPIDDSLDYGGFAHMLGDFSPQMLELMRLVYVTHSDHEGG	240

```
|||||
QY 181 OVPVAAVYVRMYKNGQVPLDSDYCGNFAHMLGFDSPQMLRLVYVTHSDHEG 240
Db 241 NVSAHTCHLVGSPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVWDECGENISTEOL 300
QY 241 NVSAHTCHLVGSPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVWDECGENISTEOL 300
Db 301 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 360
QY 301 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 360
Db 361 LELGKVKNPWPNVDAHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLPLER 420
QY 361 LELGKVKNPWPNVDAHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLPLER 420
Db 421 PKSVTMEWLEKFKRRRA 437
QY 421 PKSVTMEWLEKFKRRRA 437
RESULT 2 PRELIMINARY; PRT: 472 AA.
ID O80433;
AC O80433;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN DCCS
OS DAUCUS CAROTA (CARROT).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; ARALIALES; APIACEAE; DAUCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. MS YONSUN;
RA TAKITA E., KOYAMA H., SHIRANO Y., SHIBATA D., HARA T.;
RT *CONA encoding carrot mitochondrial citrate synthase.*;
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL; AB017159; D1033521;
DR PROSITE; PS00480; CITRATE SYNTHASE; 1.
KW LYASE: TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 472 AA; 52656 MW; D46C38CC CRC32;
Query Match 88.4%; Score 2867; DB 10; Length 472;
Best Local Similarity 85.4%; Pred. No. 0.00e+00;
Matches 373; Conservative 42; Mismatches 22; Indels 0; Gaps 0;
Db 36 ASDLDLRSQLELIPEOQERIKKKAHKGKVLQGNITVDWVIGMGMGTGLLWETSLLDP 95
QY 1 SSNDLRSQLELIPEOQERIKKKEFGSFLQGNINVDWVIGMGMGTGLLWETSLLDP 50
Db 96 FEGIPFGLSLPECKLLPAKGGEPPLPEGLLWLLTGKVPKEQVDALSALRAAV 155
QY 61 EEGIRFPGFSIPECKLLPAASAGAEPPLPEGLLWLLTGKVPKEQVDALSADLPKRA 120
Db 156 PEHYVKTIDALPITAHMPTQFATGVNALQVSEFQKAYEKGIHKTYSWEPTEDSITL 215
QY 121 PDHYVKTIDALPITAHMPTQFCTGVNALQTRSEFQKAYEKGIHKSKEWPTEDCLSL 180
Db 216 OLPVAAVYVRMYKNGQSITDSDYGANFAHMLGYSQSMOELMRLVYVTHSDHEG 275
QY 181 OVPVAAVYVRMYKNGQVPLDSDYCGNFAHMLGFDSPQMLRLVYVTHSDHEG 240
Db 276 NVSAHTCHLVASALSDPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVWSECGENISTEOL 335
QY 241 NVSAHTCHLVGSPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVWDECGENISTEOL 300
Db 301 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 360
QY 241 NVSAHTCHLVGSPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVWDECGENISTEOL 300
```

```
|||||
Db 336 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSNLFEVWPPIL 395
QY 301 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 360
Db 396 TELGKVKNPWPNVDAHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLPLER 455
QY 361 LELGKVKNPWPNVDAHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLPLER 420
Db 456 PKSVTMEWLEHCKSS 472
QY 421 PKSVTMEWLEKFKRRRA 437
RESULT 3 PRELIMINARY; PRT: 474 AA.
ID O64869;
AC O64869;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN F411.16
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL; AC004521; G3128180;
DR PROSITE; PS00480; CITRATE SYNTHASE; 1.
KW LYASE: TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 474 AA; 52782 MW; 64FF2AC6 CRC32;
Query Match 87.4%; Score 2834; DB 10; Length 474;
Best Local Similarity 85.3%; Pred. No. 0.00e+00;
Matches 371; Conservative 39; Mismatches 25; Indels 0; Gaps 0;
Db 36 STDLDLRSQLELIPEOQDRLLKLLKSEHGKVLQGNITVDWVIGMGMGTGLLWETSLLDP 95
QY 1 SSNDLRSQLELIPEOQERIKKKEFGSFLQGNINVDWVIGMGMGTGLLWETSLLDP 60
Db 96 EEGIRFGLSLPECKVLPATGAAGAEPLPEGLLWLLTGKVPKEQVEALSOLDANRAV 155
QY 61 EEGIRFPGFSIPECKVLPATGAAGAEPLPEGLLWLLTGKVPKEQVDALSADLPKRA 120
Db 156 PDYVNAIDALPITAHMPTQFASGVNALQVSEFQKAYEKGIHKSKEWPTEDCLNIA 215
QY 121 PDHYVKTIDALPITAHMPTQFCTGVNALQTRSEFQKAYEKGIHKSKEWPTEDCLSL 180
Db 216 RVPVAAVYVRMYKNGDSIPSDKSLDYGANFSLMGFDDEKVKELMRLYVTHSDHEG 275
QY 181 QVPVAAVYVRMYKNGQVPLDSDYCGNFAHMLGFDSPQMLRLVYVTHSDHEG 240
Db 276 NVSAHTCHLVGSAISDPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVVEEGEDISKEOL 335
QY 241 NVSAHTCHLVGSPDLYSFAAALNGLAGPLHGLANQEVLLMIKSVVDEGENISTEOL 300
Db 336 KEYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 395
QY 301 KDYVWKTLSNGKVVPGFGLGVLRKTDPRYTCOREFALKHLPDDPFQOLVSKLYEVVPPIL 360
Db 396 TELGKVKNPWPNVDAHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGLALER 455
```

```

QY 361 LEIGKVNPNVNDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGGLPLER 420
Db 456 PKSVTMDWLEAHCKK 470
      |||||:|||| ||:
QY 421 PKSVTMEWLEKFCR 435

RESULT 4
ID Q24135 PRELIMINARY: PRT: 469 AA.
AC Q24135;
DT 01-JAN-1998 (TREMELREL. 05, CREATED)
DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN CIT1.
OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
OC EUPHYLLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS;
OC ASTERIDAE: SOLANANAE; SOLANALES; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, SNN;
RX LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: X84226; E137432; -.
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE: TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 469 AA; 52492 MW; 6BE3C325 CRC32;

Query Match 86.3%; Score 2800; DB 10; Length 469;
Best Local Similarity 85.7%; Pred. No. 0.00e+00;
Matches 373; Conservative 34; Mismatches 27; Indels 1; Gaps 1;

Db 35 SSGDLRSELQELIPQOQDLRLKKIKSEHGKVGQGNITVDVLMGMRGTMGLLWETSLLDP 94
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 SSNDLRSELQELIPQOQDLRLKKIKSEFGSFLQGNINVDVLMGMRGTMGLLWETSLLDP 60

Db 95 DEGIRFRGLSIPECQKVLPAKPGGPELPEGLLWLLTGKVPKSEKQVDSLSQELRSRATV 154
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 EEGIRFRGFSIPECQKLLPAASAGAEPLEGLLWLLTGKVPKSEKQVDSLSQELRSRASI 120

Db 155 PDHYKTTIDALPTAHPTQFATGVMAQVQSEFQAYEKGIHKSLEPTVEDSMSLIA 214
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 PDHYKTTIDALPTAHPTQFCTGVMAQVQSEFQAYEKGIHKSLEPTVEDSMSLIA 180

Db 215 QVPLVAAYVYRMKNGNTIPKDDSLDYGANFAHMLGFSDDHMLKLYVTIHSDEGG 274
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 QVPVAAYVYRMKNGQVPLDDSLDYGANFAHMLGFSDDHMLKLYVTIHSDEGG 240

Db 275 NVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVEECGENISKEQL 334
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 NVSAHTGHLVSGPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVWDECGENISTEOL 300

Db 335 KDVAWTKLSGKVPVFGHGVLRKTPRYTCQREFALKHLPEDPFLQIVAKLYEYFQLP 394
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 KDVAWTKLSGKVPVFGHGVLRKTPRYTCQREFALKHLPEDPFLQIVAKLYEYVVPIL 360

Db 395 QNLAKL-NPWPVNDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGGLPLER 453
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 361 LEIGKVNPNVNDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGGLPLER 420

Db 454 PKSVTMEWLEHCKK 468
      |||||:||||: ||:
QY 421 PKSVTMEWLEKFCR 435

```

```

RESULT 5
ID Q43175 PRELIMINARY: PRT: 471 AA.
AC Q43175;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMELREL. 09, LAST ANNOTATION UPDATE)
DE ETHANOLAMINE AMONIA-LYASE (EC 4.3.1.7).
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
OC EUPHYLLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS;
OC ASTERIDAE: SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, DESIREE;
RX MEDLINE: 96073431.
RA LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RL "Mitochondrial citrate synthase from potato: predominant expression
in mature leaves and young flower buds.";
RL PLANTA 196:756-764(1995).
CC -1- CATALYTIC ACTIVITY: ETHANOLAMINE = ACETALDEHYDE + NH(3).
CC -1- COFACTOR: COBALAMIN.
DR EMBL: X75082; G483510; -.
DR PFAM: PF00285; citrate_synt; 1.
DR MENDEL; 15200; SOLcc;2977;mn15200.
KW LYASE.
SQ SEQUENCE 471 AA; 52612 MW; BC96A65C CRC32;

Query Match 81.3%; Score 2637; DB 10; Length 471;
Best Local Similarity 81.3%; Pred. No. 0.00e+00;
Matches 356; Conservative 41; Mismatches 36; Indels 5; Gaps 5;

Db 35 SSGDLRSELQELIPQOQDLRLKKIKSEHGKVGQGNITVDVLMGMRGTMGLLWETSLLDP 93
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 1 SSNDLRSELQELIPQOQDLRLKKIKSEFGSFLQGNINVDVLMGMRGTMGLLWETSLLDP 59

Db 94 PDEGIRFRGLSIPECQKVLPAKPGGPELPEGLLWLLTGKVPKSEKQVDSLSQELRSRASI 153
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 60 PEEGIRFRGFSIPECQKLLPAASAGAEPLEGLLWLLTGKVPKSEKQVDSLSQELRSRASI 119

Db 154 ISLIMYTTIDALPTAHPTQFATGVMAQVQSEFQAYEKGIHKSLEPTVEDSMNL 213
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 120 IPDHV-YKTIDALPTAHPTQFCTGVMAQVQSEFQAYEKGIHKSLEPTVEDSMNL 178

Db 214 IAQVPLVAAYVYRMKNGNTIPKDDSLDYGANFAHMLGFSDDHMLKLYVTIHSDEH 273
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 179 IAQVPLVAAYVYRMKNGQVPLDDSLDYGANFAHMLGFSDDHMLKLYVTIHSDEH 237

Db 274 EGGNVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVEECGENISK 333
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 238 EGGNVSAHTGHLVSGPLSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVWDECGENIST 297

Db 334 EQLKDYVWKTLSNGKVPVFGHGVLRKTPRYTCQREFALKHLPEDPFLQIVAKLYEYFVL 393
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 298 EQLKDYVWKTLSNGKVPVFGHGVLRKTPRYTCQREFALKHLPEDPFLQIVAKLYEYVVP 357

Db 394 LFLQNLAKL-NPWPVNDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGGLP 452
      :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 358 PILLELGKVNPNVNDHSGVLLNHYGLTEARYYTVLFGVSRSLGICSQLIWDRLGGLP 417

Db 453 LERPKSVTMEWLENOCKK 470
      |||||:||||:||||: ||:
QY 418 LERPKSVTMEWLEKFCR 435

RESULT 6
ID Q24259 PRELIMINARY: PRT: 339 AA.
AC Q24259;
DT 01-JAN-1998 (TREMELREL. 05, CREATED)
DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE)
DE (OXALOACETATE TRANSACETASE).

```

GN CIT1
OS POPULUS DELTOIDES X POPULUS BALSAMIFERA SUBSP. TRICHOCARPA.
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMERYPHYTA: TRACHEOPHYTA;
OC EUPHYLLIPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS: ROSIDAE;
OC VITALES: SALICACEAE: POPULUS.
RN [1]
RP SEQUENCE FROM N.A.
RA LA COGNATA U., LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER R.;
PL PLANT CELL PHYSIOL. 0-0-0(0)
CC -/- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
DR ENBL, X84227, E276838, ..
DP PFAM: PF00285; citrate_synt; 1
KW LYASE.
SQ SEQUENCE 339 AA; 37834 MW; B2DA24A9 CRC32;

Query Match 66.8%; Score 2166; DB 10; Length 339;
Best Local Similarity 85.4%; Pred. No. 0.00e+00;
Matches 286; Conservative 29; Mismatches 20; Indels 0; Gaps 0;

Db 1 MPTKEQVGLSKELRDLRALVHDYVFKALDLPVTAHPMTQFATGVMALQVQSEFQKAYEK 60
QY 101 VPSKEQVDAISADLRKRAISPDHYVTIDALPITAHPTQFCTGVMALQTRSEFQKAYEK 160
Db 61 GIHKSKEWPTYEDSLIAARPIVASIYRIYKDGKVPIMNDSLVIGNFSHMLGFDS 120
QY 161 GIHKSKEWPTYEDSLIAQVPAAYVRYRMKNGQVPLDDSLDYGNGFAHMLGFDS 220
Db 121 PEMOELMRLVYTHNSHEGNGVSAHTGHLVASALSADPYLSFAAALNGLAGPLHGLANQEV 180
QY 221 POMLEMLRLVYTHSHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLANQEV 280
Db 181 LLIWKSVEECGENTTEQLKDYVWKVTLNSGKVPVPGFGLVLPKTPRYTCOREFALKHL 240
QY 281 LLIWKSVEECGENTTEQLKDYVWKVTLNSGKVPVPGFGLVLPKTPRYTCOREFALKHL 340
Db 241 PDDPLFOLVSKLXEVVPPVLTOLGKVKPNPNVDASHGVLNLYGLTEARYTYVLFVGSV 300
QY 341 PDDPLFOLVSKLXEVVPPVLTOLGKVKPNPNVDASHGVLNLYGLTEARYTYVLFVGSV 400
Db 301 SIGICSLIWDALGLPLERPKSVTMELENHCKK 335
QY 401 SLGICSLIWDALGLPLERPKSVTMELENHCKK 435

RESULT 7 PRELIMINARY; PRT; 466 AA.
ID O75390
AC O75390
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA: PRIMATES;
OC CATARRHINI: HOMINIDAE: HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA GOLDENTHAL M.J.;
RT Cloning and sequence analysis of human citrate synthase cDNA.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -/- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -/- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -/- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -/- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR ENBL: AF047042; G3288815; ..
DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
KW LYASE, TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 466 AA; 51706 MW; C0532604 CRC32;

Query Match 63.9%; Score 2072; DB 4; Length 466;
Best Local Similarity 83.9%; Score 270; DB 1; Length 372;
Matches 272; Conservative 75; Mismatches 90; Indels 1; Gaps 1;

Best Local Similarity 62.1%; Pred. No. 0.00e+00;
Matches 272; Conservative 75; Mismatches 90; Indels 1; Gaps 1;

Db 28 ASSNLKDLADLIPKEQAPIKTFROQHGTIVGVQITVDMYGMGMKGLVFTSVLDP 87
QY 1 SSNLDLRSELQELIPEQOERLKKIKKEFGSFGQFNGVLMVGLGMRGTMGLLWETSLLDP 60
Db 88 DEGRIFRCSFPECKLLPKAKGGEELPEGLFWLLVTGCIPTTEQVSWLSKEWAKBAAL 147
QY 61 BEGRIFRGSFPECKLLPAASGAEPPEGLLWLLTGKVPSEFQVDAISADLPKRASI 120
Db 148 PSHVVTMLDNFPTNLHFMSSQSAATVATNLSNESFAQAYARGISPTKYWELIYEDVOLIA 207
QY 121 PDHYVTIDALPITAHPTQFCTGVMALOTRSEFQKAYEKGIHKSKEWPTYEDCLSLIA 180
Db 208 KLPVAAKIYRNLYWEGSGIGAIIDNSNDWHSNFMGLYTDHQFTLMRLYLIITHSDEG 267
QY 181 QVPVVAAYVYRMKNGQVI-PLDDSLDYGNGFAHMLGFDSQMLMRLYVYTHSDEG 239
Db 268 GNVSARTSHLVGSALSADPYLSFAAAMNGLAGPLHGLANQEVLYWLTOLQKEVGDVSDK 327
QY 240 GNVSARTSHLVGSALSADPYLSFAAAMNGLAGPLHGLANQEVLYWLTOLQKEVGDVSDK 299
Db 328 LRDIWTNLNSGRVPGVGHAVLRKTPRYTCOREFALKHLNDPMPFKLVAQLYKIPNV 387
QY 300 LKDYVWKVTLNSGKVPVPGFGLVLPKTPRYTCOREFALKHLNDPMPFKLVAQLYKIPNV 359
Db 388 LLEOGKAKPNPNVDASHGVLNLYGLTEARYTYVLFVGSALQVLAQLIWSRALGFPLE 447
QY 360 LLELGKVKPNPNVDASHGVLNLYGLTEARYTYVLFVGSALQVLAQLIWSRALGFPLE 419
Db 448 RPKSMSTEGMLKMFVDSKS 465
QY 420 RPKSVTMELENHCKK 437

RESULT 8 PRELIMINARY; PRT; 372 AA.
ID O28929
AC O28929;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (CITZ).
GN AF1340.
OS ARCHAEoglobus FULGIDUS.
OC ARCHAEA: EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae;
OC ARCHAEoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.F.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KEPLAVAGE A.R., GRAHAM D.E., KYPPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., PEICH C.I., MCNEIL L.K., BADGER J.H., GLOCKER A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL NATURE 390:364-370(1997).
DR ENBL: AE001011; G2649234; ..
DR TIGR: AF1340; ..
DR PFAM: PF00285; citrate_synt; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 372 AA; 42270 MW; DB9662BC CRC32;

Query Match 8.3%; Score 270; DB 1; Length 372;
Best Local Similarity 27.2%; Pred. No. 8.07e-31;


```

RESULT 14
ID O54382 PRELIMINARY; PRT; 356 AA.
AC O54382;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PUERTO RICO;
RX MEDLINE: 98087556.
RA DAVIS M.J., YING Z., BRUNNER B.R., PANTOJA A., FERWERDA F.H.;
RT "Rickettsial relative associated with papaya Bunchy top disease.";
RL CURR. MICROBIOL. 36:80-84(1998).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL; U76908; G2894797; -.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1 356
FT NON_TER 356 356
SQ SEQUENCE 356 AA; 40030 MW; 15BF5D0 CRC32;

Query Match 6.9%; Score 223; DB 2; Length 356;
Best Local Similarity 28.7%; Pred. No. 1,55e-21;
Matches 72; Conservative 69; Mismatches 86; Indels 24; Gaps 19;

Db 108 FKEADYELTAIRMIKIPTIAAMSYK--YSIGQPFYVDPNSLDFTENFLHM-FATPCEK 164
QY 167 FWEPTYE-DCLSLIAQVPVAAVYVRRMYKNGOVIP-LDSDLYGNGFAHMLGFDSP-QM 223
Db 165 YKVPVKNALNKIFLHADHEQ-NASTSTVRIAGSGANPFACVSTGIASLWGPAGGA 223
QY 224 --LE-LMR--LY-VTI-HSDHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHGLA 276
Db 224 NEAVINMLKDIGSVENIPKYIAKAKDNFRLMGFGRHVKNYD-P--RAAVLKETCKE 280
QY 277 NQEVLLWIKSV--VDECGENISTEQLKDYVWKTLSNG-KVVPFGGLGVLRKTDPRYTQCR 333
Db 281 VLKELGOLDNNPLLOIAIEL-BAIA-LKDEYFIERKLYPNVDYFSGIYKAMGIP-PQMF 337
QY 334 EF-ALKHLPDDPFFOLVSKLYEVVPPILLELGKVKKNPWPNDVDAHSGVLLNHYGLTEARY 392
Db 338 TVLFATARTVG 348
QY 393 TVLFGVSRSLG 403

RESULT 15
ID Q59777 PRELIMINARY; PRT; 411 AA.
AC Q59777;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HA-91;
RA RAOULT D., RYDKINA E., POUX V., BREMEEVA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

```

```

CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL; U59731; G1390011; -.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1 411
FT NON_TER 411 411
SQ SEQUENCE 411 AA; 46162 MW; 9921533A CRC32;

Query Match 6.9%; Score 223; DB 2; Length 411;
Best Local Similarity 29.0%; Pred. No. 1,55e-21;
Matches 73; Conservative 67; Mismatches 88; Indels 24; Gaps 20;

Db 150 NFKADYELTAIRMIKIPTIAAMSYK--YSIGQPFYVDPNSLDFTENFLHM-FATPCT 206
QY 166 KFEPTYE-DCLSLIAQVPVAAVYVRRMYKNGOV-IPLDSDLYGNGFAHMLGFDSP-- 221
Db 207 KYTVNPIKNALNKIFLHADHEQ-NASTSTVRIAGSGANPFACISTGIASLWGPAGG 265
QY 222 Q-MLE-LMR--LY-VTI-HSDHEGNGVSAHTGHLVGSPLSDPYLSFAAALNGLAGPLHL 275
Db 266 ANEAVINMLKEIGSEVPIKYIAKAKDNFRLMGFGRHVKNYD-P--RAAVLKETCK 322
QY 276 ANQEVLLWIKSV-VDEC-GENISTEQLKDYVWKTLSNG-KVVPFGGLGVLRKTDPRYTQ 332
Db 323 EVLKELGOLDNNPLLOIAIEL-BAIA-LKDEYFIERKLYPNVDYFSGIYKAMGIP-SQM 379
QY 333 REF-ALKHLPDDPFFOLVSKLYEVVPPILLELGKVKKNPWPNDVDAHSGVLLNHYGLTEARY 391
Db 380 FTVLFATARTVG 391
QY 392 YTVLFGVSRSLG 403

Search completed: Fri Oct 22 16:54:47 1999
Job time : 151 secs.

```

W O R L D
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Oct 23 20:18:09 1999; MasPar time 3158.85 Seconds
Tabular output not generated. 1533.004 Million cell updates/sec

Title: >US-08-702-718-5
Description: (1-1747) from US08702718.seq
Perfect Score: 1747
N.A. Sequence: 1 GCTCTTGGGATCTATTTCCT.....AAAAA.....TTTTTTTTTTTTTTTT
Comp: CGAGAACCTAGATAAGGA.....TTTTTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158

1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_ov 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_v1

Database: genbank11
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 12.066; Variance 12.847; scale 0.939

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1747	100.0	1747	25	A46547 Sequence 3 from Patent	0.00e+00
2	1747	100.0	1747	27	Nicotiana tabacum mRNA for cit	0.00e+00
3	1238	70.9	1891	25	A46545 Sequence 1 from Patent	0.00e+00
4	1238	70.9	1891	27	STCITS S. tuberosum mRNA for m	0.00e+00
5	848	48.5	1859	27	ABO17159 Daucus carota mRNA for	0.00e+00
6	818	46.8	1733	27	CMU19481 Citrus maxima citrate	0.00e+00
7	726	41.6	1851	25	A46546 Sequence 2 from Patent	2.39e+268
8	726	41.6	1851	27	BVCITSYNH B. vulgaris mRNA for ci	2.29e+268
9	655	37.5	1680	27	ATWCTISYNH Arabidopsis thaliana m	7.23e+240
10	596	34.1	1378	27	PHCITSYNH Populus hybrid mRNA fo	3.18e+216
11	274	15.7	1401	31	AF047042 Homo sapiens citrate s	2.34e+88
12	273	15.6	1455	23	PICITSYNH Porcine citrate synth	5.77e+88
13	210	12.0	1481	27	SCWTCISNA S. cerevisiae mitochond	1.94e-63

14	210	12.0	2427	27	SCCS01	Yeast gene for citrate	1.94e-63
15	210	12.0	4391	27	SCYNR001C	S.cerevisiae chromosom	1.94e-63
16	210	12.0	23901	27	SCN201952	S.cerevisiae N2019, N2	1.94e-63
17	194	11.1	1800	27	AB001565	Candida tropicalis DNA	2.88e-57
18	163	9.3	24643	27	SPAC6C3	S.pombe chromosome I c	2.13e-45
19	148	8.5	5430	27	SC1	Yeast (S.cerevisiae) C	1.04e-39
20	148	8.5	315339	27	SCCHRII1	S.cerevisiae chromosom	1.04e-39
21	146	8.4	1680	27	YSCCIT2	Yeast (S.cerevisiae) C	5.91e-39
22	131	7.5	1710	27	ASNCIT1	Aspergillus niger mRNA	2.56e-33
23	114	6.5	2638	27	NEUCITIA	Kluyveromyces lactis d	5.30e-21
24	99	5.7	562	34	KLAJ9626	T. thermophila mRNA fo	1.65e-21
25	87	5.0	1497	21	TETCITSYN	Cenorhabditis elegans	3.46e-17
26	87	5.0	48096	21	CET20G5	human STS Wt-12459	5.34e-14
27	76	4.5	467	34	GI3212	Cenorhabditis elegans	2.68e-13
28	76	4.4	152184	19	CEY70G10	Arabidopsis thaliana c	5.99e-13
29	75	4.3	118327	28	AYAC004521	Emricella nidulans ci	8.16e-07
30	57	3.3	2398	28	ENU89675	S.cerevisiae citrate s	7.44e-05
31	51	2.9	2790	27	SCCIT3SYN	S.cerevisiae chromosom	7.44e-05
32	51	2.9	17660	27	SC9723	Saccharomyces cerevisi	7.44e-05
33	51	2.9	38779	27	SCU31900	S.cerevisiae chromosom	7.44e-05
34	51	2.9	165536	27	SCCHRXVI	Sequence 14 from paten	6.76e-04
35	48	2.7	7218	25	I66494	Sequence 17 from Paten	2.89e-03
36	47	2.7	7218	25	I66494	S.oleracea L. mRNA for	2.46e-02
37	46	2.6	3437	25	A52184	P.falciiparum mRNA for	2.46e-02
38	46	2.6	3437	25	SOPULSPO	Mus musculus transcrip	2.46e-02
39	43	2.5	1920	21	PFGTUB	Drosophila melanogaste	2.46e-02
40	43	2.5	3600	32	MUSAMYB	Plasmodium falciparum	2.46e-02
41	43	2.5	10772	21	AF032089	B.verrucosa Betv III m	4.96e-02
42	43	2.5	202291	19	AC004153	Dictyostelium discoide	4.96e-02
43	42	2.4	762	27	BVBETVII1	Dictyostelium discoide	4.96e-02
44	42	2.4	1840	21	DDU73686		
45	42	2.4	4875	21	AB009080		

ALIGNMENTS

RESULT 1
LOCUS A46547 1747 bp DNA
DEFINITION Sequence 3 from Patent WO9524487.
ACCESSION A46547
NID Q2300711
VERSION A46547.1 GI:2300711

KEYWORDS common tobacco.

ORGANISM

Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Solanales; Solanales; Solanales; Solanales; Solanales;
1 (bases 1 to 1747)

REFERENCE

AUTHORS Mueller-Roeber, B., Landschuetz, V. and La, C.U.
TITLE PROCESSES FOR INHIBITING AND FOR INDUCING FLOWER FORMATION IN PLANTS

JOURNAL

Patent: WO 9524487-A 3 14-SEP-1995;
HOECHST SCHERING AGREVO GMBH (DE)
Other publication DE 4438821 960425
Other publication DE 4435366 960328
Other publication AU 2067995 950925
Other publication DE 4408629 950914.

FEATURES

Location/Qualifiers
1..1747
/organism="Nicotiana tabacum"
/db_xref="taxon:4097"
/tissue_type="LEAF"
/clone="TCS"
70..1479
/SC_number="4.1.3.7"
/codon_start=1
/product="CITRATE SYNTHASE"
/protein_id="CAA02910.1"
/db_xref="PID:e306253"
/db_xref="PID:g2300712"
/db_xref="GI:2300712"
/translation="MVFYRGVSLKLSRAVQNTNLSNVFWLQVOTSSGLDRSEL"

QELIPQODRLKLLKSEHKKVYOLGNITVDMVLGMRGMTGLLWETLSLDDPDEIRFG
LSIYECQKVLPAKGGEPLEGLLWLLTGTGKVPSEKQVDSLSQELSRATVPBHVX
TIDALPVTAKMNGTIPATGMALQVSEFOKAYERGIHKSUKLWEPTIYEDMSLIAQVPL
VAYYRRMYKNGTIPKDDSLDYGANFAHMLGSSSDHMLKLYYIHSDEHGGNV
SAHTGLHVASLSDPYLSFAALNGLAGPLHLGAEVLLWIKSVFEEGGENI SKEQL
KDYAMKTLKSGVVPFGHGLVRKTPRYTCQREFKALHPEDLPFLVAKLYEVLQ
FLONLAKLPWPNVDAHSGVLLNYYGLTEARYIYVLFVGSRALGICSLIWDRLGLP
LERPSYMEWLENHCKKA*
BASE COUNT 490 a 335 c 400 g 522 t
ORIGIN

Query Match 100.0%; Score 1747; DB 25; Length 1747;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GCTCTTGGGATCTATTCCTCTCTCTATTTCTCCCTAGTAGTAAAGTTAATTTGGTATT 60
Qy 1 GCTCTTGGGATCTATTCCTCTCTCTATTTCTCCCTAGTAGTAAAGTTAATTTGGTATT 60

Db 61 TTGGGAGCCATGGTCTATCGCGGGTTCTCTGCTCAAGCTGCGTCTCGAGG 120
Qy 61 TTGGGAGCCATGGTCTATCGCGGGTTCTCTGCTCAAGCTGCGTCTCGAGG 120

Db 121 GTCCACAGACAAATCTTAGCAACTCTGTGGGTGGCTTCAAGTCCAAACCTCTTCTGT 180
Qy 121 GTCCACAGACAAATCTTAGCAACTCTGTGGGTGGCTTCAAGTCCAAACCTCTTCTGT 180

Db 181 CTGTGATCTGGTCTGAGCTGCAGAAATGTATCCAGAACACAGAGTGCCTTAAGAG 240
Qy 181 CTGTGATCTGGTCTGAGCTGCAGAAATGTATCCAGAACACAGAGTGCCTTAAGAG 240

Db 241 CTCAGTCAAGCATGGAAGGTTCATTTGGGAAACATCACAGTTGATAGTTCTTGT 300
Qy 241 CTCAGTCAAGCATGGAAGGTTCATTTGGGAAACATCACAGTTGATAGTTCTTGT 300

Db 301 GGAATGAGAGGAATGACAGGATTAATTTGGGAAACCTTACTTTGACCCGATGAAGA 360
Qy 301 GGAATGAGAGGAATGACAGGATTAATTTGGGAAACCTTACTTTGACCCGATGAAGA 360

Db 361 ATTCGCTTTCGGGCTTGCTATCTATGAATGCCAAAGGTATACCTGCAGCAAGCT 420
Qy 361 ATTCGCTTTCGGGCTTGCTATCTATGAATGCCAAAGGTATACCTGCAGCAAGCT 420

Db 421 GGGGAGAGCCCTTGCTGAGGCTCTCTGCTTCTTTTAAAGGAAAGGTGCGCATCA 480
Qy 421 GGGGAGAGCCCTTGCTGAGGCTCTCTGCTTCTTTTAAAGGAAAGGTGCGCATCA 480

Db 481 AAGAGCAAGTGAATCTATGCTCAGGAATTCGGAAGTCGTGCTACTGCCCGATCAT 540
Qy 481 AAGAGCAAGTGAATCTATGCTCAGGAATTCGGAAGTCGTGCTACTGCCCGATCAT 540

Db 541 GTATCAAAACTATGATGCTTACCAGTCACAGCTCATCCAAATGACTCAGTTTGTACT 600
Qy 541 GTATCAAAACTATGATGCTTACCAGTCACAGCTCATCCAAATGACTCAGTTTGTACT 600

Db 601 GGAGTCATGCTCTTCAGGTTCAAAGTGAATTTCAAAGGCGATATGAGAAGGGATTTCAC 660
Qy 601 GGAGTCATGCTCTTCAGGTTCAAAGTGAATTTCAAAGGCGATATGAGAAGGGATTTCAC 660

Db 661 AATCAAAAGTTATGGAAACCGACATATGAGATTCCATGATTTGATGCTCAAGTTCCA 720
Qy 661 AATCAAAAGTTATGGAAACCGACATATGAGATTCCATGATTTGATGCTCAAGTTCCA 720

Db 721 CTGTGCTGCTTATGTTTATCCAGGATGTACAAGACGGCAACACTATACCTAAGAT 780
Qy 721 CTGTGCTGCTTATGTTTATCCAGGATGTACAAGACGGCAACACTATACCTAAGAT 780

Db 781 GACTCACTGGAATATGGTGCAAATTTTGTGCATGCTTGGTTTCAGTAGCTCTGACATG 840
Qy 781 GACTCACTGGAATATGGTGCAAATTTTGTGCATGCTTGGTTTCAGTAGCTCTGACATG 840

Db 841 CATGAGCTTATGAGCTCTATGTCAGATACACAGTGCATCAAGTGTGTAACCTCAGT 900
Qy 841 CATGAGCTTATGAGCTCTATGTCAGATACACAGTGCATCAAGTGTGTAACCTCAGT 900

Qy 841 CATGAGCTTATGAGCTCTATGTCAGATACACAGTGCATCAAGTGTGTAACCTCAGT 900
Db 901 GCTCACAGGTCACTTGGTTGCTAGTGTCTTTGTACAGCCCTTACCTCTCTCGTGTCT 960
Qy 901 GCTCACAGGTCACTTGGTTGCTAGTGTCTTTGTACAGCCCTTACCTCTCTCGTGTCT 960
Db 961 GCTTTGAATGGTTAGCTGACCACTTTCATGGTTTACCAATCAGGAAGTTTTCATGG 1020
Qy 961 GCTTTGAATGGTTAGCTGACCACTTTCATGGTTTACCAATCAGGAAGTTTTCATGG 1020

Db 1021 ATCAAACTCTGTTAGAGAGTGTGGGAGAACATTTCCAAAGAGCAGTTGAAAGACTAC 1080
Qy 1021 ATCAAACTCTGTTAGAGAGTGTGGGAGAACATTTCCAAAGAGCAGTTGAAAGACTAC 1080

Db 1081 GCTTGGAAACATTTGAAAGTGGCAAGTGTCTCCCTGGTTTCGAGACATGGAGTTCTGCGC 1140
Qy 1081 GCTTGGAAACATTTGAAAGTGGCAAGTGTCTCCCTGGTTTCGAGACATGGAGTTCTGCGC 1140

Db 1141 AAGACTGATCAAGATACACATGCCAGAGAGTTCCTGCTTGAAGCATTTGCCCTGAAGAT 1200
Qy 1141 AAGACTGATCAAGATACACATGCCAGAGAGTTCCTGCTTGAAGCATTTGCCCTGAAGAT 1200

Db 1201 CCAGTGTTCACACTGTTGCAAACTCTAGCAAGTGTCTCCCTCAATTTCTTACAGAACTTG 1260
Qy 1201 CCAGTGTTCACACTGTTGCAAACTCTAGCAAGTGTCTCCCTCAATTTCTTACAGAACTTG 1260

Db 1261 GCAAAGTTAAACCTTGGCCAAATGTTGATGCCACAGTGTGTGTTGTTGAACTATTAT 1320
Qy 1261 GCAAAGTTAAACCTTGGCCAAATGTTGATGCCACAGTGTGTGTTGTTGAACTATTAT 1320

Db 1321 GGTTTAACTGAAGCAAGATATATACGGTCTCTTTGGTGTATCAAGAGCTCTTGGCATT 1380
Qy 1321 GGTTTAACTGAAGCAAGATATATACGGTCTCTTTGGTGTATCAAGAGCTCTTGGCATT 1380

Db 1381 TGCTCTCAGCTAATTTGGGACCGAGCTCTTGGATTCGCTAGAGAGGCCAAAGAGTGC 1440
Qy 1381 TGCTCTCAGCTAATTTGGGACCGAGCTCTTGGATTCGCTAGAGAGGCCAAAGAGTGC 1440

Db 1441 ACAATGAGTGGCTTGAGAACCATTCGAAGAACAGTATTTGTTGAAATCTCTCGGAG 1500
Qy 1441 ACAATGAGTGGCTTGAGAACCATTCGAAGAACAGTATTTGTTGAAATCTCTCGGAG 1500

Db 1501 CATAAAAGCAATGTAATTTATGAATAATTTGTTGAGAAAGCAGTTTTTCTTGG 1560
Qy 1501 CATAAAAGCAATGTAATTTATGAATAATTTGTTGAGAAAGCAGTTTTTCTTGG 1560

Db 1561 AGCCAAGTAGTGGCATTAGGATGTTTCATCGATTGGCTTAGTACGGTTTTTGAAGATT 1620
Qy 1561 AGCCAAGTAGTGGCATTAGGATGTTTCATCGATTGGCTTAGTACGGTTTTTGAAGATT 1620

Db 1621 TGGTTGTGATTTTCAGTTCGGTTTTTAAATGTTTATACCAATACCTTATCGATATAA 1680
Qy 1621 TGGTTGTGATTTTCAGTTCGGTTTTTAAATGTTTATACCAATACCTTATCGATATAA 1680

Db 1681 TTCAATATGATTCGATTTTACTTTTGTGAAAAAACAACAAAAAATAAAAAA 1740
Qy 1681 TTCAATATGATTCGATTTTACTTTTGTGAAAAAACAACAAAAAATAAAAAA 1740

Db 1741 AAAAAA 1747
Qy 1741 AAAAAA 1747

RESULT 2 NTCITSYNH 1747 bp mRNA PLN 07-OCT-1996
LOCUS N. tabacum mRNA for citrate synthase.
DEFINITION X84226
ACCESSION g1556428
NID X84226.1 GI:1556428
VERSION cit1 gene; citrate synthase.
KEYWORDS common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

FEATURES		Location/Qualifiers	
source		1..1859	
		/organism="Daucus carota"	
		/cultivar="MS Yonsun"	
		/db_xref="taxon:4039"	
		/cell_type="suspension cultured cells"	
5'UTR		1..156	
gene		157..1575	
CDS		157..1575	
		/gene="DCCS"	
		/gene="DCCS"	
		/EC_number="4.1.3.7"	
		/codon_start=1	
		/product="citrate synthase"	
		/protein_id="BA32557.1"	
		/db_xref="PID:d1033521"	
		/db_xref="PID:g3493367"	
		/db_xref="GI:3493367"	
		/translation="MVFFRSVLLNKLRSVQSNLNTVRWFQVOTSSDLDLRSQ LKLIPEQQRKIKLKAEGKVLGNITVDMLGMRGTMGLLWETSLDPEGIREF GLSIPCOKLLPGAKGPELPGLGLWLLTGKPTKEQVDALSAELSRPAVPEHY KTDLAPVTAHPMTQATGVMALQVSEFOKAYEKIHKTKYWEPTSDITLIAQLP VVAAYIYRMKNGQSTSDSDIDYGANFAHMLGYDPSMQLMRLYVTIHTDEGNN VSATGHLVASALSDPYLFAAALNGLAPLHGLANQVLLWIKSVVSCGENVTKQ LKDYINWTLNSGRVPGYGHVGLRNDTPRYICQREFALKHLPDPLQVLSNLFVVP PILTELKVKNPWPNDVHSGVLLNHVGLTEARYTVLFGVSAIGICSQLVDRLAG LPLERKSVTMEWLENHCKSS"	
3'UTR		1576..1859	
polyA_signal		1838..1843	
polyA_site		1859	
BASE COUNT		521 a 363 c 412 g 563 t	
ORIGIN			
Query Match		48.5%; Score 848; DB 27; Length 1859;	
Best Local Similarity		80.7%; Pred.No. 0.00e+00;	
Matches		1155; Conservative 0; Mismatches 271; Indels 6; Gaps 3;	
Db	152	GAGAAATGGTCTTCTCGGAGTGTTCGCTGCTTAACAAGCTCGGCTCGGCTGTC	211
Qy	65	GAGCATGGTCTTCTATCGCGCGCTTCTCTGCTGTCAAAGCTCGTCTCGACGCGTCC	124
Db	212	AGCAGTGAACCTCAGTAATACAGTGGTGGTTTCAAGTGCAGACCTCGCTCTCTATC	271
Qy	125	AACAGCAAAATCTTAGCAACTCTGTGGGCTTCAAGTCCAAACCTC---TTCTGGTC	181
Db	272	TCGATCTTCGTTCTCAACTTAAGGAGGTGATTCCAGAACAGCAGGAGGATTAAAAAC	331
Qy	182	TGATCTCGTCTGAGCTGCAAGATTGATCCAGAACACAGAGATCGCTAAAGAGC	241
Db	332	TAAAGCAGAACATGGAAAGCTTCAGCTGGGAAACATAACTTGTATATGTTAGTTCGCG	391
Qy	242	TCAAGTCAGAGCATGGAAGGTTCAATTGGGAAACATCAGATTGATGTTCTTGGTG	301
Db	392	GTATGAGAGGATGACAGGACTTCTATGGGAAACCTCTTTACTTGACCCAGAGAGGAA	451
Qy	302	GAATGAGAGGAATGACAGGATTACTGTGGGAAACCTCATTACTTGACCCGATGAAGAA	361
Db	452	TTCGTTTAGGGGTGTGCCATACCTGAATGTCAAGAGCTATTACCAGAGCGAAGCCG	511
Qy	362	TTCGTTTCGGGCTGTCTATCTATGATGCCAAAGGATTACTTGACCAAGGCTG	421
Db	512	GTGAGAACCATTCGCTGAGGCTGCTGCTGGCTTCTTTTAACGTGGAAGGTACCACTA	571
Qy	422	GGGAGAGCCCTTCCTCGAAGGCTTCTCTGGCTCTTTTAAACAGGAAGGTGCATCAA	481
Db	572	AAGAGAGTAGATGCAATGTCCGACAGTTCGCAAGTTCGCTGCTGTACCAGAGCATG	631
Qy	482	AAGAGAGTAGATGCAATGTCTCAGGAATTCGCAAGTTCGCTGCTGTACCAGAGCATG	541
Db	632	TATCAAAACCATCGATCGCTTACTCTTACAGCTCATCCCAATGACTCAATTTGCAACTG	691
Qy	542	TATCAAAACCATCGATCGCTTACTCTTACAGCTCATCCCAATGACTCAATTTGCTACTG	601

Db 61 CTGACAGCATTAAGAAAGAAATTTGGAGTTTCCAGCTGGGGAATATCAATGTTGACATG 120
QY 232 CTAAGAAGCTCAAGTCAGAGCATGGAAGGTTCAATTTGGGAACATCACAGTTGATATG 291
Db 121 GTATTGGCGGAATGAGAGGAATGACTGTTTACCTTTTGGGAGACTTCCTTACTCGACCCA 180
QY 292 GTTCTGGTGAATGAGAGGAATGACAGGATTAATGTTGGGAACCTCATTTACTTGACCCC 351
Db 181 GAAGAGGATTCGGTTTCAGGGGTTTTTCTATACCTGAAATGCGGAAACTTTTACCCGCT 240
QY 352 GATGAAGGAATTCGGTTTCGGGGTGTCTATCTATGAATGCCAAAGAGTATTAACCTGCA 411
Db 241 GCAAGTGTCTGCGAGGCAATTCCTGAAGGTCTTCTTGGCTTCTTTTAAACGGGAAG 300
QY 412 GCAGAGCTTGGGGAGAGCCCTTGGCTGAGGTCTTCTGCTGCTTTTAAACAGGAAG 471
Db 301 GTTCTAGCAAGAGCAAGTAGACTCTATCAGCAGATTTACGAAACGCTTCTTATC 360
QY 472 GTGCCATCAAAAGAGCAAGTGATTCATTGCTCAGGAATTCGGAAGTCGTACTGTC 531
Db 361 CCAGACCATGTGTACAAACAATTTGATGCTCTACCTATTTACGGCTTCCATGACTGAC 420
QY 532 CCGGATCATGTATACAAACTATTGATGCTTACCAGTCACAGCTCATCCATGACTGAC 591
Db 421 TTTTGCATCGGTGTTATGCGCTTACAGACTCGAAGCGAATTTCAAGAGGCATATGAGAA 480
QY 592 TTTGCTACTGGAGTCATGCTCTTACAGTTCAAGTGAATTTCAAGAGGCATATGAGAA 651
Db 481 GGGATCCATTAAGTCAAGTTTTGGGAGCAACATATGAGGAGTGCCTTAGTTGATGCT 540
QY 652 GGGATTCACAAATCAAAATTTAGGGAACCGACATATGAGGATTCATGATTTGATGCT 711
Db 541 CAGATTCCTGTTGTCAGCTTATGTTATCGGAGGATGTAAGAAATGGCAAGTATA 600
QY 712 CAGATTCCTGTTGTCAGCTTATGTTATCGGAGGATGTAAGAAATGGCAAGTATA 600
Db 601 CCCTGATGATGCTTCCCTTATGATGTTGGAATTTTGGCACACATGTTGGGATTTGATAGC 660
QY 772 CCTAAGGATGACTCACTGATGATGTTGTCGAATTTTGTTCACATGCTTGGTTTCACTAGC 831
Db 661 CCTCAGATGCTGAGCTGATGGCTTTTATGTCACAAATTCACAGTATGATGAGGTTGA 720
QY 832 TCTGACATCATGAGCTTATGAGCTTCTATGTCAGGATACAGAGTATGATGAGGTTG 891
Db 721 AATGTTAGTGCACACTGGCCATTTGGTGGTGTAGTCCACTTTCAGATCCTTATTTGCA 780
QY 892 AAGTCAAGTGCACACAGCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 951
Db 781 TTTGACGACCATTAATGTTGCTGGCCACTCCATGATGATGATGATGATGATGATGATGATG 840
QY 952 TTTGCTGCTGCTTGAATGTTTGTAGTGGACCACTTTCATGTTTGTAGTGGTGGTGGTGGT 1011
Db 841 CTGCTGCTGATTAATGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
QY 1012 TTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1071
Db 901 AAGATATTGTTTGGAGACACTTAACAGTGGCAAGGTTTGTACCTGGATTTGCTCTAGGA 960
QY 1072 AAGATACGCTTGGAAACATTTGAAAGTGGCAAGGTTTGTCCCTGGTTCGGACATGA 1131
Db 961 GTATTGGGAAGACATCCAGATACATGCGCAAGAGATTTGCGTTTGAAGCAGCTTG 1020
QY 1132 GTTCTGGCAAGACTGATCCAGATACATGCGCAAGAGAGTTCGCTTTGAAGCAGCTTG 1191
Db 1021 CCGTATGACCCATTTTCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1080
QY 1192 CCGTATGACCCATTTTCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1250
Db 1081 TTAGAGCTTGGAAAGGTAAGAAATCCATGCGCTTAATTTGATGCTCATAGTGGAGTTTG 1140
QY 1251 ACAGAACTTGGCAAGTTAA--ACCTTGGCCAAATTTGATGATGATGATGATGATGATGATG 1308

Db 1141 CTGAACCACTATGTTTGCAGAGCAAGATACATATACGTTTGTGTTGGGTATCAAGS 1200
QY 1309 TTGAACCTATTATGTTTAACTGAAGCAAGATATATACGCTCTCTTGTGTATCAAGA 1368
Db 1201 AGTCTTGGAAATGCTCACAGCTTATATGGACCGAGCTTCTTGGCTTGGCTAGAGAG 1260
QY 1369 GCTCTTGGCAATTTGCTCTCAGCTAATTTGGGACCGAGCTTCTGATTTGCCACTAGAGG 1428
Db 1261 CCAAAGAGTGTCACTATGGAATGGCTTGAAGATTTTGTAA 1301
QY 1429 CCAAAGAGTGTCAATTTGGAGTGGCTTGAGAACCATTTGCAA 1469
RESULT 8
LOCUS BVCITSYNH 1551 bp mRNA PLN 07-OCT-1996
DEFINITION B.vulgaris mRNA for citrate synthase.
ACCESSION X84228
NID g1556379
VERSION X84228.1 GI:1556379
KEYWORDS cit1 gene; citrate synthase.
SOURCE beet.
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
REFERENCE 1 (bases 1 to 1551)
AUTHORS La Cognata U., Landschuetz V., Willmitzer, L. and Mueller-Roeber, B.
JOURNAL Plant Cell Physiol. In press
REFERENCE 2 (bases 1 to 1551)
AUTHORS La Cognata U.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1995) U. La Cognata, Institute fuer
Genbiologische Forschung, Berlin GmbH, Ihnestr. 63, 14195
Berlin, FRG
FEATURES
source 1..1551 Location/Qualifiers
/organism="Beta vulgaris"
/db_xref="taxon:3555"
gene 1..1314
/gene="cit1"
CDS <1..1314
/gene="cit1"
/EC_number="4.1.3.7"
/codon_start=1
/product="citrate (si)-synthase"
/protein_id="CAA59010.1"
/db_xref="PID:el37433"
/db_xref="PID:g1556380"
/db_xref="GI:1556380"
/translation="SSNLDRLSELQELIPEQOERLKKIKKEFSFOLGNIYNDVVLGG
MRQMTGLLWETSLDPGRIPEQKILPAASAGAPLPGLLWLLTGVK
SREQVDAISADLRKRASIPDHVKITDALPITAHPTOFCTGMALOTRSEFQKAYEK
GIHKSKEPTEYEDCLSLAQVVAAYVRYRMKNGQVPLDDSLDYGNAHMLGF
DSQDELMLRIYTIHSDHEGNNVAHTHLVSGPLSDPLSFAALNGLAGPLHGLA
NOFVLEMLRVDECEENISBOLKDYVMKTLNSGVKVPGLGVLRKTDPRYTCORE
FALKHLDPDFQVSKLYEVPVPILELKVKNPWPNDHAHSGVLRLLNHYLPLEARY
TVLFVYSRSLGICSLIDRALGLPLERPKSVTMEWLEKFKRRA"
BASE COUNT 455 a 268 c 370 g 458 t
ORIGIN
Query Match 41.6%; Score 726; DB 27; Length 1551;
Best Local Similarity 78.5%; Pred. No. 2.29e-268;
Matches 1021; Conservative 0; Mismatches 277; Indels 3; Gaps 2;
Db 1 TCCTCTAACCTTGACCTCGTTCAGAGTTACAGAGTTCAGAACTGATTCCTGAACACAGGACGA 60
QY 172 TCTTCTGGCTTGTGCTGCGTTCTGAGCTGCAAGATTTGATTCAGAACACAGGATCGC 231
Db 61 CTGAAGAGATTAAGAAGAAATTTGGAAGTTTCCAGCTGGGGAATATCAATGTTGACATG 120
QY 232 CTAAGAAGCTCAAGTCAGAGCATGGAAGGTTTCAATTTGGAAACATCACAGTTGATATG 291

Db	121	GTATTGGGCGGAATGAGAGGAATGACTGGTTTACTTTGGGAGACTTCCTTACTTCGACCCA	180
Qy	292	GTCTTTGGTGGAAATGAGAGCAATGACAGGATTACTGTGGGAAACCTCATTACTTGACCCC	351
Db	181	GAAGAGGGTATCCGGTTCAGGGGTTTTTCTATACCTGAATGCCAGAACTTTTACCGGT	240
Qy	352	GATGAGGAATTCGCTTTCCGGGCTTGTCATCTATGAATGCCAAAAGGATTACTGTGCA	411
Db	241	GCAAGTGCCTGGTGACAGGCATTGGCTGAAGTGCTCTCTTTGGCTTCTTTTAAACGGAAAG	300
Qy	412	GCAAAGCCCTGGGGAGAGCCCTTGCTGAAGTGCTCTCTCGCTTCTTTTAAACAGAAAG	471
Db	301	GTTCTCTGAAGAAGCAAGTAGATGCTCTATCAGCAGATTTACGAAACAGTGCTTCTATC	360
Qy	472	GTGCCATCAAAAGCAGCAAGTGGAATCATTTGCTCAGGAATTCGCAAGCTGCTGCTACTGC	531
Db	361	CCAGACCATGTGTACAAAACAATTCATGCTCTACCTATTATCGGCTCATCAATGACTCAG	420
Qy	532	CCCGATCATGTATACAAAATATTGATGGCTTACCAGTCACAGCTCATCCATGACTCAG	591
Db	421	TTTTGCATGGTGTATTGGCCTTTACAGACTCGAAGCGAATTTTCAGAAGGCATATGAGAAA	480
Qy	592	TTTGCTACTGGAGTCATGGCTCTTCAGGTTCAAAGTGAAATTTCAAAAGGCATATGAGAAA	651
Db	481	GGGATCCATAAGTCAAAGTTTTGGGAGCCACATATGAGGACTGCGCTTAGTTGATTGCT	540
Qy	652	GGGATTCACAATCAAAGTTATGGGAACCGACATATGAGGAATCCATGAGTTTGATTGCT	711
Db	541	CAAAGTCTCTGTTTGCAGCTTATGTTTTATCGAGGATGTATAAGATTGACAAGTAATA	600
Qy	712	CAAAGTCCACTTGTCTGCTTATGTTTTATCGCAGGATGTACAAGAAGCGCAACACTATA	771
Db	601	CCGCTGGATGACTCCCTTGATTTGGTGGAAATTCGCACACATGTTGGGATTTGATAGC	660
Qy	772	CCTAAGATGACTCACTGGATTTGGTGCAAATTTTGCTCACATGCTTGGTTTCAGTAGC	831
Db	661	CCTCAGATGCTTGAGCTGATGCGCCTTTATGTCACAATTCACAGTATCATGAGGTGGA	720
Qy	832	TCTGACATCATGAGCTTATGAAGTCTATGTCAAGATACACAGTATCATGAAGTGCT	891
Db	721	AATGTTAGTGCACACACTGGCCATTTGGTGGTAGTCCACTTTCAGATCCTTATTGTCA	780
Qy	892	AAGCTCAGTGCTCACACAGGTCACCTGGTGTGTAGTGCTTGTGAGACCCCTTACCCTCC	951
Db	781	TTTGCAGCAGCATTAATGGTTTGGCTGGCCACTCCATGGATAGCCAAACACAGGAATC	840
Qy	952	TTGCTGCTGCTTTGGAATGGTTTTAGCTGGACCACTTCATGGTTTTAGCCAATCAGGAAT	1011
Db	841	CTGCTGTGANTTAATCAGTTGTTGATGAATGTGGAGAGAACATCTCGACAGACAGTTG	900
Qy	1012	TTGCTATGATCAAAATCTGCTAGCAGAGTGTGGGAGAACATTTCCAAGAGCAGTTG	1071
Db	901	AAAGATTATGTTTGAAGACACTAAACAGTGGCAAGTTGTACCTGGATTGGTCTAGGA	960
Qy	1072	AAAGACTACGCTTGGAAACATTTGAAGTGGCAAGTTGTCCCTGGTTTCGGNACATGGA	1131
Db	961	GTATTCGGGAAGACAGATCCAAGATACACATGCCAAAGAGAAATTTGGGTTGAAGCACTG	1020
Qy	1132	GTCTCGGAAGACTGATCCAAGATACACATGCCAGAGAGAGTTGCGTTTGAAGCAATTTG	1191
Db	1021	CTGTATTGACCCATTTTTCAAATTTGGTGTCAAAGTTGTATGAAGTGGTCCCTCATTTCTA	1080
Qy	1192	CCTGAAGATCCACTGTTTCACTGGTTCGAAACTCTACGAAGTGTTTCTTCNATTCCT	1250
Db	1081	TTAGACTTGGAAAGTAAAGAAATCCATGGCCTAATTTGATGCTCATATGGAGTTTG	1140
Qy	1251	ACAGAACTTGGCAAGTTAA--ACCCCTTGGCCAAATGTTGATGCCACAGTGGTGTTG	1308
Db	1141	CTCAACCACCTATGGTTTGACAGCAAGACATACATACGGTTTTTGTGGGTATCAAGG	1200
Qy	1309	TTGAACATAATATGGTTTAACTGAACAGATATTATACGGTCTCTTTGGGTATCAAGA	1368
Db	1201	AGTCTTGGAAATATGCTACACAGCTTATATGGGACCAAGCTCTTGGCTTGGCGCTAGAGG	1260

QY	1369	GCTCTTGGCAATTGCTCTCAGCTAATTTGGACCGAGCTCTTGTGATTGCCACTAGAGAGG	1428
Db	1261	CCAAGAGTGTCATCTAGGAATGCGTTGAAAGTTTGTAA	1301
QY	1429	CCAAGAGTGTCATCTAGGAATGCGTTGAGAACCTTGCAA	1469
RESULT	9		
LOCUS	ATMTCITRN	1680 bp	PLN
DEFINITION	Arabidopsis thaliana mRNA for mitochondrial citrate synthetase.		
ACCESSION	X17528		
NID	g11243		
VERSION	X17528.1	GI:11243	
KEYWORDS	citrate synthase.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.		
AUTHORS	Unger, E.A., Hand, J.M., Cashmore, A.P. and Vasconcelos, A.C		
TITLE	Isolation of a cDNA encoding mitochondrial citrate synthase from Arabidopsis thaliana		
JOURNAL	Plant Mol. Biol.	13 (4), 411-418	(1989)
MEDLINE	91370823		
FEATURES	Location/Qualifiers		
source	1..1680		
	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
CDS	58..1476		
	/EC_number="4.1.3.7"		
	/codon_start=1		
	/product="citrate synthetase"		
	/protein_id="CAA35570.1"		
	/db_xref="PID:e1188578"		
	/db_xref="PID:g2652924"		
	/db_xref="GI:2652924"		
	/db_xref="SWISS-PROT:P20115"		
	/translation="MVFFERSVSAFTRLSRPVGGQSSLSNVWIKQWSSDTDLKSLQLELPEHDLRLKLSKSEHGKVGOLNITVDVVGMRGTGLLWETSLDPEVFAFGDRLCPKQALLPTAGSGGUNHFRSFVAALNWKTKLAKSKLHKRKTWNRRAVSDIARYNADALPTAHPTMTQFASGVMAQVQSEFKAYENGHKSKEWPTYEDCLNLIARVPPVAAYVRRMYKNGDSIPDSKGLDYGANFSHMLGDFDERLKEMLRITSPETVMHGGNVSAAETHGLVGSALSDPSISFAAALNGLAGHLGANOELLWKSVVECGDLSKEGLFELSEEPWNVDAHSGLVLIHNYGLTEARYTYLVFGVSRSLGICSQLLDRELLIALEPRKSVTMDWLEAHCKKASSA"		
BASE COUNT	439 a	345 c	488 t
ORIGIN			
Query Match	37.5%	Score 655;	DB 27; Length 1680;
Best Local Similarity	76.9%	Pred. No. 7.23e-240;	
Matches	1057; Conservative	0; Mismatches 300;	Indels 17; Gaps 16;
Db	111	GGGCGACAAATCTTCACTCAGCAATTCGTGAGATGAGTTCAGATGCAGAGCTCTACCGA	170
QY	120	GSTCCACACAGACAATCTTAGCAACTCTGTGCGGTGCGTTCAGTCCAAACCTCTCTGG	179
Db	171	CCTGACCTGAAGTCGAGCTGCAAGAGTTAATTCGCGAACACAGAGACCTCTGAAGAA	230
QY	180	TCTTGATCTGGTTCTGAGCTGCAAGATTCGATTCAGAACACAGAGTCCCTTAAGAA	239
Db	231	ACTGAAGTCAGACATGGGAAGGTCCAACTGGGAACATCACTGTTGATATGGTTATGG	290
QY	240	GCTCAAGTCAGACATGGGAAGGTTCAATTGGGAACATCACAGTTGATATGGTTCTTGG	299
Db	291	TGGAATGAGGGGATGACTGGATTGCTTTGGAAACCTCATTTGTTGACCCGGAAGA	349
QY	300	TGGAATGAGGGATGACAGGATCTACTGTGGAAACCTCATTTGTTGACCCGATGAGG	359
Db	350	TATTCGCTTTAGGGGATGTGCACTTCTGTGAGTCCCAAGAACCCCTATTACCTACTGCC	409

```
QY 360 AATTGCTTTTCGGGCTTGTCTATCTA-TGAATGCCA-AAAGG--TATTACCTGCGAGCAA 415
Db 410 AGTCTGGAGGCTTGAACCAATACCGAGGCTCTTTTGTGGCTTCTCTTAACCTGGAAGGTA 469
QY 416 AGCCTGGGAGAGC-CCTTGCCTGAAGGTCTCTCTGCTCTCTTTTAACAGGAAGGTTG 474
Db 470 CTTAGCAAGAAGCAAGTTGAAGCACTGTGAAGAGCTTG-GAAC-CGTGCTGCTGTGTC 527
QY 475 CCATCA-AAAGAGCAAGTGGATTCTCTCAGSAATTGCGAAGTCTGCTACTGTCCTCC 533
Db 528 AGATTATGTGTAACAATGCCATGCTGCTCTGCTCTCCACAGCTCATCCATGACTCAATT 587
QY 534 CGATCATGTATACAAACTATTGATGCTTACCAGTACAGCTCATCCATGACTCAATT 593
Db 588 TGTACGGGTGTATGCGGCTCCAGGTGCAAGTGAAGTTTCAAAGGCATATGAGAAATG 647
QY 594 TGTACTGGAGTCACTGCTTTCAGGTTCAAAGTGAATTTCAAAGGCATATGAGAAAG 653
Db 648 AATTCATAAGTCAAAAGTTCTGGAGCCCAACATATCAGGATTGCTCAACCTGATTGCTCG 707
QY 654 GATTCAAAATCAAGTTATGGAACCGACATATGAGGATTCCATGAGTTGATTGCTCA 713
Db 708 TGTCTGTTGTAGCTGCATATGTTTATCGAAGGATGTATGAAGATGCTGATTCATTC 767
QY 714 AGTTCCACTTGTGCTGCTTATGTTTATCGCAGGATGTACAAGAGCGCAACACTATCC 773
Db 768 CTCAGATAAATCTTGGATTATGGTGCAAAATTTTCCACATGTATGATGATGATGA 827
QY 774 TAAGGATGACTCACTGGATTTATGGTGCAAAATTTTCTCACATGCTGTGTTTCACTAG 833
Db 828 AAGGTTTGAAGAGCTCATGAGCT-TACATACCACTCACAGTGTATGATCAAGGTGAA 885
QY 834 TGACATGCTAGCTTATGAAGCTCTATGTCACGATACACAGTGTATCATGAAGTGTGA 892
Db 887 ATGTTAGTGTACACATGCTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 946
QY 893 ACGTGAGTGTCTACAGAGTCACTTGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 952
Db 947 TTCAGCTGATTAATGTTTGTAGCTGGGCACTCCATGTTTGGCTTAATCAGGAAGTTT 1006
QY 953 TCCTGCTGCTTGAATGTTTGTAGCTGGACACTTCACTGTTTACCAATCAGGAAGTTT 1012
Db 1007 TGCTTTGGATCAATCAGTCTGAGAGGAATGTGGAAGAGATATATCAAAAGAACAGTTGA 1066
QY 1013 TGCTATGATCAATCTGTTGAGAGGAGTGTGGGAGAGAACATTTCCAAAGAGCAGTTGA 1072
Db 1067 AAGNATATGTTTGAAGAACATTAACAGTGGCAAGTGTATTCGGGATATGTCAGCTG 1126
QY 1073 AAGACTACGCTTGGAAACATTTGAAAGTGGCAAGTGTGCTGCTGTTGCGACATGGAG 1132
Db 1127 TTCTGCGCAATCTGATCCAAAGATATGTATGCAAGAGAAATTTGCTTTGAAGCATCATC 1186
QY 1133 TTCTGCGCAAGCTGATCCAAAGATACACATGCCAGAGAGTGTGCTTTGAAGCATTTGC 1192
Db 1187 CTGACGACCTCT-TTTCAGTG-TTGCAAG-CT-TATGAAGTTGGCTCTCTCT-CTCAC 1241
QY 1193 CTGAAGATCCACTGTTTCAACTGGTTGCAAACTCTACGAAGTGTTCCTCCAAATTCCTAC 1252
Db 1242 TGAGCTTGAAGTGAAGAACCTT-GGCCAAATGTTGATGCTCACAGTGGGGTCTTGTCTGA 1300
QY 1253 AGAAGTGGCAAGATTAAACCTTGGCCAAATGTTGATGCCACAGTGGTGTGTTGTTGA 1312
Db 1301 ACCACTATGCTTAACCGAAGCAAGTACTACACCGTGTCTTGGTGTGTTTCAAGAGCTC 1360
QY 1313 ACTATTATGTTTAACTGAAGCAAGATATATACGGTCTCTTGGTGTATCAAGAGCTC 1372
Db 1361 TTGCATCTGCTCTCACTTATATGGACCGAGAGCTCTTGTGACATTGAGAGGCCAA 1420
QY 1373 TTGCGATTGCTCTCACTTAATTGGGACCGAGCTCTTGGATTGCGCCTAGAGAGGCCAA 1432
Db 1421 AGAGTGTATACCATGGACTGCTTGAAGCCCAATTTGAAGAAAGCTTCATCTGCTT 1474
QY 1433 AGAGTGTACAAATGGAGTGGCTTGAGNACCAATTTGCAAGNACGATGATTTGTTT 1486
```

```
PHCITSYNH 1378 bp mRNA PLN 28-OCT-1996
Populus hybrid mRNA for citrate synthase.
X84227
g1556430
X84227.1 GI:1556430
citi gene: citrate synthase.
Populus balsamifera subsp. trichocarpa X Populus deltoides.
Populus balsamifera subsp. trichocarpa X Populus deltoides
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Violales; Salicaceae; Populus.
1 (bases 1 to 1378)
La Cognata,U., Landschuetze,V., Willmitzer,L. and Mueller-Roeber,B.
Plant Cell Physiol. In press
2 (bases 1 to 1378)
La Cognata,U.
Direct Submission
Submitted (23-JAN-1995) U. La Cognata, Institute fuer
Genbiologische Forschung, Berlin GmbH, Imnestrasse 63, 14195
Berlin, FRG
Location/Qualifiers
1..1378
/organism="Populus balsamifera subsp. trichocarpa X
Populus deltoides"
/note="hybrid"
/db_xref="taxon:3695"
16..1035
/gene="citi"
16..1035
/gene="citi"
/EC-number="4.1.3.7"
/codon_start=1
/trans_except-(pos:16..18,aa:Met)
/product="citrate (si)-synthase"
/protein_id="CAA59009.1"
/db_xref="PID:e276838"
/db_xref="PID:g1648926"
/db_xref="GI:1648926"
/db_xref="SPTRMBL:O24259"
/translation="MPTKEQVGLSKELDRALVHYVFKALDALPVAHPMTQFAIG
VMAIQVSEFOKAEKGIHKSQWEPYEDSLIARVPIVASITIRRIKDKQKIPM
NDSLVIGNFSEHMLFDFSPQEMRLVYTNHSDHEGNNVSAHTHLVASALSDPYLS
FAAALNGLAGPLHGLANQEVLLTKSVVEECENITEQLKDYVTKLNSGVVPGFG
HGVLRKTVPRVTCOREFALKHLPDDPLQVLSKLYEVVPPVLQGLKRNKPNVDAH
SSVLLNYGLTEARYTVLFGVSRISIGCSQLIWDRLGLPLERPKSVTMELLENHCK
KAPAN"
BASE COUNT 396 a 252 c 313 g 417 t
ORIGIN
Query Match 34.1%; Score 596; DB 27; Length 1378;
Best Local Similarity 80.3%; Pred. No. 3.18e-216;
Matches 810; Conservative 0; Mismatches 196; Indels 3; Gaps 2;
Db 14 AGGTACACCAAGAGAGAGTGGTGGCTTATCCAGGAATTCGCTGACCGTCTGCTGG 73
QY 470 AGGTGCGCATCAAAGAGCAAGTGGATTCTTCTCAGGAATTCGCAAGTCTGCTACTG 529
Db 74 TTTCATGATTATGCTTTTAAGGCACTTGTGCTTACCTGCTGCTCATCCATGACTC 133
QY 530 TCCCGCATGATATACAAACTATTGATGCTTACCAGTCACAGCTCATCCATGACTC 589
Db 134 AGTTTCAACTGGTGTATGGCCCTACAGGTTACAGAGTGAATTTCAAAGACCTATGAA 193
QY 590 AGTTTCTACTGAGTCATGGCTCTTCAGGTTCAAAGTGAATTTCAAAGGCATATGAGA 649
Db 194 AAGGATTATCAATAAGTACTGGAGCCCAACATATGAGGATTCCTTAGTCTGATTG 253
QY 650 AAGGGATTCACAATCAAAAGTTATGGGAACCCACATATGAGGATTCATGAGTTGATTG 709
Db 254 CTCGTGTGCCAATAGTAGCTTCATATATTTATCGAAGGATTTTACAAGATGGGAAGTTA 313
```


Db 1084 CAAGAGAGTTTCTCTGTAACACCTTGCCTCATGACCCCATGTTTAAGCTGTTGCTCAG 1143
 Qy 1165 CAGAGAGAGTTTCTCTGTAACACCTTGCCTCATGACCCCATGTTTAAGCTGTTGCTCAG 1224
 Db 1144 CTGTACAAAGTTTCTCTGTAACACCTTGCCTCATGACCCCATGTTTAAGCTGTTGCTCAG 1203
 Qy 1225 CTGTACAAAGTTTCTCTGTAACACCTTGCCTCATGACCCCATGTTTAAGCTGTTGCTCAG 1281
 Db 1204 AATGTGAGTCTCAGAGTGGGGTCTGCTCCAGTACTATGGCAGTACGAGGAGTAAGTAC 1263
 Qy 1282 AATGTTGATGCCACACAGTGGTGTGTTGTTGAATATATGTTTAACTGAAGCAAGATAT 1341
 Db 1264 TACACAGTCTGTCGGGTATACAGGGCACTGGGTGTTATGACACAGTCACTGAGGC 1323
 Qy 1342 TATACGGTCTCTTGGTGTATCAAGAGCTCTTGGCATTTGCTCAGCTAATTTGGGAC 1401
 Db 1324 CGAGCCTTAGCTTCCCTCTAGAGAGGCCCAAG 1356
 Qy 1402 CGAGCTCTTGATGCCACTAGAGAGGCCCAAG 1434

RESULT 13
 LOCUS SCMTCS1NA 1481 bp DNA PLN 17-FEB-1997
 DEFINITION S.cerevisiae mitochondrial citrate synthase gene, complete CDS.
 ACCESSION 223259
 NID 9313749
 VERSION 223259.1 GI:313749
 KEYWORDS mitochondrial citrate synthase.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 1481)
 AUTHORS Lindner,P.
 JOURNAL Thesis (1993) Protein Engineering, Max-Planck-Institut fuer Biochemie
 REFERENCE 2 (bases 1 to 1481)
 AUTHORS Lindner,P and Plueckthun,A
 TITLE The effect of point mutations in the hinge of yeast citrate synthase
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1481)
 AUTHORS Lindner,P.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUL-1993) Peter Lindner, Protein Engineering, Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, Martinsried, D-82152, Fed. Rep., Germany
 REFERENCE 4 (bases 1 to 1481)
 AUTHORS Suissa,M., Suda,K. and Schatz,G.
 TITLE Isolation of the nuclear yeast genes for citrate synthase and fifteen other mitochondrial proteins by a new screening method
 JOURNAL EMBO J. 3 (8), 1773-1781 (1984)
 MEDLINE 85003587
 REMARK (sites)
 FEATURES
 source Location/Qualifiers
 1..1481
 /organism="Saccharomyces cerevisiae"
 /strain="D273-10B"
 /db_xref="taxon:4932"
 /clone="Plasmid pFCS1 (Suissa et al., EMBO J. 3, 1773-1781 (1984))."
 /gene="CIT1"
 /db_xref="SGD:L0000341"
 1..1481
 sig_peptide 1..111
 /gene="CIT1"
 /db_xref="SGD:L0000341"
 /evidence="experimental"
 /product="the processing direct N-terminal protein sequencing"
 gene 1..1481
 /gene="CIT1"
 CDS 1..1440
 /gene="CIT1"

/codon_start=1
 /db_xref="SGD:L0000341"
 /product="mitochondrial citrate synthase"
 /protein_id="CAA80781.1"
 /db_xref="PID:g313750"
 /db_xref="GI:313750"
 /db_xref="SWISS-PROT:P00890"
 /translation="MSALLSTSKSFSLRSRSTFQCONMUKALFALLNAKHYSASRQT
 LKRFABIIPAKAEIKKFKHGKTVIGEVLLQAYGGMGRGKGLVWEGSVLDPEEG
 IFRGRITPEIQRELPAKGESEPLPEALFWLLTGEIPTDAQVKALSADLAARSEIP
 EHVLIQDLSLPKDLHPMAQFSAIATVALESEKFAKAYAGCKEYWSYTFEDSLDL
 GKLPVLASKIYRNVFQDKKITSTDPNADYGNLAOLLYGKDFIDLMRLYLTIHSDH
 EGNVSAHTTHLVGSLSPYLSLAAGNLGAGLHGRANOEVLWFLKLRFEVKGDY
 SKETIKYLDWTLNAGRVVPGYGHAVLRKTDPRYTAQREALKHFPDYELFKLVSTLY
 EVAPGLIKHGKTNFNPVNDSHGVLQYIGLTAEASFYIVLFGVARAIGVLPOLLID
 RAYGAPLIERKSTFKYKELVKKIESKN"
 172
 conflict
 /gene="CIT1"
 /citation=[4]
 /db_xref="SGD:L0000341"
 /replace="g"
 234..235
 conflict
 /gene="CIT1"
 /citation=[4]
 /db_xref="SGD:L0000341"
 /replace="a"
 1472
 conflict
 /gene="CIT1"
 /citation=[4]
 /db_xref="SGD:L0000341"
 /replace="a"
 452 a 266 c 319 g 444 t
 BASE COUNT
 ORIGIN
 Query Match 12.0%; Score 210; DB 27; Length 1481;
 Best Local Similarity 60.3%; Pred. No. 1.94e-63;
 Matches 688; Conservative 0; Mismatches 448; Indels 5; Gaps 4;
 Db 243 TGGTGGTATGAGAGGTATTAAAGGCTTTGTTGGGAAGGTTTCGCTGTAGACCCCGAAGA 302
 Qy 297 TGGTGGTATGAGAGGATGACAGGATTACTGTGGAAACCTCATCTTACCCCGGATGA 356
 Db 303 AGGTATTAGATTTAGGGGTGCTACTATTCCAGAAATTCAAAGGGAACCTCAAAAGGCTGA 362
 Qy 357 AGGAATTCGCTTCGGGGCTTGTCTATCTATGAATGCCAAAGGTATTACCTGCAGCAAA 416
 Db 363 GGGTAGTACAGAACTTTGCCAGAGCTTTATTTTGGTTCGCTTTGACTGGTGAATACC 422
 Qy 417 GCCTGGGGGAGAGGCTTTCCTGGAAGGCTTCTCTGCTTTTAAACAGGAAAGGTGCC 476
 Db 423 TACTGACCTCAAGTTAAAGCCCTTTCGCTGATTATTAGTGCACAGATCAGAAATTCAGA 482
 Qy 477 ATCAAAGAGCAAGTGGATTCAATGCTCAGGAATTCGGAAGTCTGCTACTGTCCCGA 536
 Db 483 GCAGCTTATCAACTTTTAGTAGCTCCCAAAAGATCTACATCCAATGGGGCAATTTTC 542
 Qy 537 TCATGTATACAAACTATTGATGCTTACCAGTCACAGCTCATCCAATGACTCAGTTGC 596
 Db 543 TATTCCCGTACTGCTTTAGAAAGCGAGTCTAAGTTTTCGCAAGCATATGCTCAAGTGT 602
 Qy 597 TACTGGAGTCATGCTCTTCAGGTTCAAGTGAATTTCAAAGGCGATATGAGAAGGAT 656
 Db 603 ATCCAAGAAAGATATTGGAGCTATACATTTGAAGATTTCGTTAGATCTCGTGGGTAATT 662
 Qy 657 TCACAAATCAAAGTTATGGGAACCGACATATGAGGATTCATGAGTTTGATTGCTCAAGT 716
 Db 663 ACCTGTTATTGCTTCCAAAATTTATCGTAATGCTTTCAGAGGATGGTAAATTTACTTCAAC 722
 Qy 717 TCCACTTGTCTGCTTATGTTTATCGCAGGATGTACAAGAACGGCACTATACCTAA 776
 Db 723 CGATCCTAATGCTGACTATGTTAAATTTGGCCCAACTTTTGGGCTACGAAAAACAAGGA 782
 Qy 777 GGATGACTCACTGGATTATGTCGCAAAATTTTGCCTCACATGCTTGGTTTCAGTACCTCTGA 836

Db 1460 TGACAAAACGATGCTCGTTATACGGCTCAACGCTGAATTCGATTAACACATTTCCACAG 1519
 QY 1136 TGGCAAGAGTATCCCAAGATACATGCCAGAGAGTTCGCTTTGAAGCATTTCCCGT 1195
 Db 1520 ATTACGAGTTATTAACTTGGTCTCCACCAATTTATGAAGTTGCCCGAGGGTTTTAACTA 1579
 QY 1196 AAGATCCACTGTTTCACTGGTTGCAAACTTACGAAGT-GTTCCTCCAAATCTTACAG 1254
 Db 1580 AGCATGTAATAAGAACCCATGGCCAAATGTTGATTCACATTCGGTGTTTTATTCG 1639
 QY 1255 AACTTGGCAAGTTAA-ACCCTTGGCCAAATGTTGATGCCACAGTGGTGTTGTTGA 1312
 Db 1640 AATACTATGCTTAACCTAGAGCTTCGTTCTACACTGATTTGTTGTTGTTGCCAGAGCTA 1699
 QY 1313 ACTATTATGTTTAACTGAAGCAAGATATTATACGGTCCCTTTTGGTGATCAAGAGCTC 1372
 Db 1700 TTGGTGTGTACCCCAATTAATCATCATAGAGGCTGTGGTGCTCCAAATCGAAAGGCCAA 1759
 QY 1373 TTGGCATTTGCTCTACGCTAATTGGGACGAGCTCTTGATTTGCCACTAGAGAGGCCAA 1432
 Db 1760 A 1760
 QY 1433 A 1433

RESULT 15
 LOCUS SCYNR001C 4391 bp DNA PLN 11-AUG-1997
 DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNR001c.
 ACCESSION 271616 Y13139
 NID g1302468
 VERSION 271616.1 GI:1302468
 KEYWORDS baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 4391)
 AUTHORS Aert,R., Verhasselt,P., Voet,M. and Volckaert,G.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4391)
 AUTHORS MIPS.
 TITLE Direct Submission
 JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the European yeast chromosome XIV sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mews@mips.emblnet.org
 FEATURES
 source
 1. .4391
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="XIV"
 misc_feature 142..251
 /note="CEN14"
 misc_feature 142..149
 /note="CEN element CDE I"
 misc_feature 150..234
 /note="CEN element CDE II"
 misc_feature 235..259
 /note="CEN element CDE III"
 gene complement(1006..2445)
 /gene="CIT1"
 /db_xref="SGD:L0000341"
 complement(1006..2445)
 /gene="CIT1"
 /note="ORF YNR001c"
 /codon_start=1
 /db_xref="SGD:L0000341"
 /protein_id="CAA96277.1"
 /db_xref="PID:e239778"
 /db_xref="PID:g1302469"
 /db_xref="GI:1302469"
 /db_xref="SWISS-PROT:P00890"

/translation="MSALLSTKTSFLSRGSTRQCONMOKALFALLNABHYSSAQT
 LKRPAPLIPAKAEELKKFKERKGVIGVLEQLQAYGMRGKGLWEGSVLDPEEG
 IRRGRTIPEIQRELKPAEGSTPELPEALFWLLTGEITPDQVKAALSADARSEIP
 EHVILDSLPKDLHPMAQFSIAVTALESKFAKAYAGQVSKWYSVTFEDSLDL
 GKLPVASKIYRNVFKDKITSDPNADYGNKLAQLLGVENKDFIDLMRLYLTHSDH
 EGNVSAHTTHVLSALSSPYLSLAAGNLGLAGRANOEVLEWLFKLREVKGY
 SKETIEKLDWTLNAGRVVPGYGHVLRKTDPRYTAOREFALKHFPDYELFKLVSTIY
 EVAPGLTKHGKTKNPNPVDVSHSVLLQYGLTEASFYTVLFGVARAIGVLQQLIID
 RAVGAPLERPKSFSTEKYRELVRKKIESKN"
 complement(3230..3301)
 /gene="tp(AGG)NR - systematic name"
 /note="anticodon gene: AGG; tRNA-Pro2 - common name"
 complement(3230..3301)
 /gene="tp(AGG)NR - systematic name"
 complement(3382..3497)
 /note="delta-remnant"
 complement(3503..3873)
 /note="solo-tau"
 3983..4056
 /gene="tn(GTT)NR - systematic name"
 /note="anticodon gene: GTT; tRNA-Asn - common name"
 3983..4056
 /gene="tn(GTT)NR - systematic name"
 BASE COUNT 1412 a 803 c 784 g 1392 t
 ORIGIN
 Query Match 12.0%; Score 210; DB 27; Length 4391;
 Best Local Similarity 60.3%; Pred. No. 1.94e-63;
 Matches 688; Conservative 0; Mismatches 448; Indels 5; Gaps 4;
 Db 1064 TTTGGCCTTTTCGATTGGAGCACCAACGCCCTATCGATGATTGGGGTAAACACACCA 1123
 Cp 1433 TTTGGCCTCTCTAGTGGCAATCCAGAGCTCGGTCCCAATAGCTGAGACCAATGCCA 1374
 Db 1124 ATAGCTCTGGCAACACCAACAAATACAGTGTAGAACGAGCCTCAGTTAGACCATAGTAT 1183
 Cp 1373 AGAGCTCTTGATACCAACAGAGGACCGTATAATATCTTGCATTAAACATAATAG 1314
 Db 1184 TGCAATAAACACCGGAATGTGAATCAACATTTGGCCATGGGTTCTTAGTTTACCATGC 1243
 Cp 1313 TTCAACAACACACCACTGTGGGCATCAACATTTGGCCNAGGGTT--TAACTTGCCAAGT 1256
 Db 1244 TAGTTAAACCCCTGGGCAACTTCATAAATGGTGGAGACCAACTTAATAAATCGTAA 1303
 Cp 1255 TCTGTAAACAATTGGAGGAACA-CCTCGTAGAGTTTTCGAACACAGTGTGAACAGTGGATCT 1197
 Db 1304 TCTGGAAATGTTCAATGCGAATTCAGTTGAGCCGTATAACGAGGATCAGTTTTTCTC 1363
 Cp 1196 TCAGGCAATGCTTCAAAAGCGAACTCTCTCGGCATGTGTATCTTGGATCAGTCTTGGC 1137
 Db 1364 AAAACCCGATGGCCATAACAGGAACAACTCTCCCTGGGTTCAAAGTATCCCAAGTAC 1423
 Cp 1136 AGAAGTCCATGTCGAAACAGGAGGACAACTTGGCACTTTTCAATGTTTCCAGCGTAG 1077
 Db 1424 T-TTTCAATTTGTTTGTGTAATAGTCACCTTTTCACTTCTTCTCTCAATTTAAATAGCCA 1482
 Cp 1076 TCTTTCAACTGCTCTTTGGAAATGTTCTCCACACACTCTCTACAACAGATTG-ATCCA 1018
 Db 1483 TCTCAAAACTTCTTGATTGGCAGCAGCATGTAATGGGCGACCTAACCATTCAACACAGC 1542
 Cp 1017 TAGCAAAACTTCTCGATTGGGCTAAACCATGAAGTGGTGGTCCAGCTAAACGACAGC 958
 Db 1543 GCGCAAGATAGTATGGCGAAGATAAGCGAGAACCCCACTAAATGTGTAGTATGGGAGA 1602
 Cp 957 AGCGAAGGAGAGGTAAGGCTCTGACAAAGCACTAGCAACCAAGTGACCTGTGTGAGCACT 898
 Db 1603 AAGCTTACACCTTCATCATCAGAAATGAATAGTTAAATATATAGTCTCATTAAGTCAATAA 1662
 Cp 897 GAGGTTACACCTTCATCATCAGTGTGTATGTCGACATAGAGCTTCATAGCTCATGAT 838
 Db 1663 ATCCCTGTTTTCGTAGCCCAAAAGTTGGGCCAAATTTTTTACCATAGTACGATAGGATC 1722
 Cp 837 GTCAGAGTACTGAACCAACCAAGCATGTGAGCAAAATTTGCACCAATATCCAGTGAGTCATC 778



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 22:01:39 1999; MasPar time 392.61 Seconds
Tabular output not generated. 953.450 Million cell updates/sec

Title: >US-08-702-718-5
Description: (1-1747) from US08702718.seq
Perfect Score: 1747
N.A. Sequence: 1 GCTCTGGGATCTATTGCT.....AAAAA.....AAAAA 1747
Comp: CGAGAACCTAGATAAGGA.....TTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.616; Variance 7.859; scale 1.224

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1747	100.0	1747	16	T04201 Tobacco citrate synth	0.00e+00
2	1238	70.9	1891	16	T04199 Potato citrate synth	0.00e+00
3	1238	70.9	1891	17	T03410 Potato citrate synth	0.00e+00
4	726	41.6	1551	16	T04300 Sugar beet citrate sy	0.00e+00
5	46	2.6	204	1	N81164 Base substituted E.co	7.45e-08
6	46	2.6	3437	21	T31478 Spinach debranching e	7.45e-08
7	43	2.5	204	1	N81164 Base substituted E.co	1.44e-06
8	43	2.5	1558	22	T28255 Survival motor neuron	1.44e-06
9	43	2.5	1550	22	T18928 Human survival motor	1.44e-06

10	43	2.5	1582	22	T28259 Survival motor neuron	1.44e-06
11	43	2.5	1582	22	T18931 Human survival motor	1.44e-06
12	43	2.5	1848	56	V84588 Human secreted protei	1.44e-06
13	44	2.5	2519	50	V54124 Human membrane protei	5.39e-07
14	40	2.3	472	56	V84530 Human secreted protei	2.64e-05
15	40	2.3	700	3	Q22352 P14(T) allergen clone	2.64e-05
16	40	2.3	700	14	Q89191 Birch pollen P14 alle	2.64e-05
17	40	2.3	1046	2	Q10377 Plasmid pMG3C9 used t	2.64e-05
18	40	2.3	1080	17	T04682 Human superoxide-dism	2.64e-05
19	41	2.3	2229	4	Q23462 Human B-raf protein k	1.01e-05
20	41	2.3	3933	16	T03865 Human mucosal lymphoc	1.01e-05
21	40	2.3	7859	7	Q44265 pSW6 for expression o	2.64e-05
22	40	2.3	7859	2	Q12154 Shuttle vector pSW6.	2.64e-05
23	40	2.3	7984	4	Q25185 pSW6 expression vecto	2.64e-05
24	38	2.2	84	41	V09269 Nucleotide sequence o	1.78e-04
25	38	2.2	85	44	V37488 Human growth regulato	1.78e-04
26	39	2.2	91	9	Q51746 Oligonucleotide probe	6.89e-05
27	39	2.2	91	9	Q51746 Oligonucleotide probe	6.89e-05
28	38	2.2	350	16	Q99380 Rat allograft inflamm	1.78e-04
29	39	2.2	581	34	T69172 Trypanosoma cruzi ant	6.89e-05
30	38	2.2	785	56	V84473 Human secreted protei	1.78e-04
31	39	2.2	880	35	T91714 Rat SmLJM/GRP2 CDNA..	6.89e-05
32	38	2.2	1118	3	Q20263 Deg-1 gene.	1.78e-04
33	38	2.2	1123	52	V34290 Human secreted protei	1.78e-04
34	38	2.2	1320	52	V34253 Human secreted protei	1.78e-04
35	39	2.2	1454	42	V23913 Plant CAD enzyme DNA	6.89e-05
36	39	2.2	1474	42	V23914 Plant CAD enzyme DNA	6.89e-05
37	39	2.2	1521	58	V73000 Human foetal brain se	6.89e-05
38	38	2.2	1694	50	V33190 Secreted protein B038	1.78e-04
39	38	2.2	1875	49	V33033 Dirofilaria immitis t	1.78e-04
40	38	2.2	1978	6	Q37189 Petunia EPSP synthase	1.78e-04
41	38	2.2	2196	51	V59663 Human secreted protei	1.78e-04
42	39	2.2	5153	21	T30347 Human YAP cDNA.	6.89e-05
43	38	2.2	5852	2	Q11710 Dictyostelium plasmid	1.78e-04
44	39	2.2	6010	1	Q11998 Human plasminogen wit	6.89e-05
45	39	2.2	6020	1	Q06648 Plasminogen gene from	6.89e-05

ALIGNMENTS

RESULT 1
ID T04201 standard; cDNA to mRNA; 1747 BP.
AC T04201;
DT 23-JAN-1996 (first entry)
DE Tobacco citrate synthase cDNA.
KW Citrate synthase; flower formation; ss.
OS Nicotiana tabacum
FH Key Location/Qualifiers
FT cds 70..1476 /*tag= a
FN WO9524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-433366.
PR 19-OCT-1994; DE-438821.
PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI: 95-328278/42.
DR P-PSDB: R82840.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Claim 7; Page 60-63; 87pp; English.
CC To identify a cDNA bank of leaf tissue from tobacco which codes for citrate
CC synthase, a cDNA bank of leaf tissue from tobacco was prep.
CC plasmids of this cDNA bank were screened using a radioactive DNA
CC probe which comprises Solanum tuberosum citrate synthase cDNA
CC (T04199). One of the clones was sequenced. The nt. sequence is
CC given in T04201.
SQ Sequence 1747 BP; 490 A; 335 C; 400 G; 522 T;
Query Match 100.0%; Score 1747; DB 16; Length 1747;

Best Local Similarity 100.0%: Pred. No. 0.00e+00:

Sequence	Length	Matches	Mismatches	Indels	Gaps
Sequence 1	1747	1747	0	0	0
Sequence 2	1747	1747	0	0	0
Sequence 3	1747	1747	0	0	0
Sequence 4	1747	1747	0	0	0
Sequence 5	1747	1747	0	0	0
Sequence 6	1747	1747	0	0	0
Sequence 7	1747	1747	0	0	0
Sequence 8	1747	1747	0	0	0
Sequence 9	1747	1747	0	0	0
Sequence 10	1747	1747	0	0	0
Sequence 11	1747	1747	0	0	0
Sequence 12	1747	1747	0	0	0
Sequence 13	1747	1747	0	0	0
Sequence 14	1747	1747	0	0	0
Sequence 15	1747	1747	0	0	0
Sequence 16	1747	1747	0	0	0
Sequence 17	1747	1747	0	0	0
Sequence 18	1747	1747	0	0	0
Sequence 19	1747	1747	0	0	0
Sequence 20	1747	1747	0	0	0
Sequence 21	1747	1747	0	0	0
Sequence 22	1747	1747	0	0	0
Sequence 23	1747	1747	0	0	0
Sequence 24	1747	1747	0	0	0
Sequence 25	1747	1747	0	0	0
Sequence 26	1747	1747	0	0	0
Sequence 27	1747	1747	0	0	0
Sequence 28	1747	1747	0	0	0
Sequence 29	1747	1747	0	0	0
Sequence 30	1747	1747	0	0	0
Sequence 31	1747	1747	0	0	0
Sequence 32	1747	1747	0	0	0
Sequence 33	1747	1747	0	0	0
Sequence 34	1747	1747	0	0	0
Sequence 35	1747	1747	0	0	0
Sequence 36	1747	1747	0	0	0
Sequence 37	1747	1747	0	0	0
Sequence 38	1747	1747	0	0	0
Sequence 39	1747	1747	0	0	0
Sequence 40	1747	1747	0	0	0
Sequence 41	1747	1747	0	0	0
Sequence 42	1747	1747	0	0	0
Sequence 43	1747	1747	0	0	0
Sequence 44	1747	1747	0	0	0
Sequence 45	1747	1747	0	0	0
Sequence 46	1747	1747	0	0	0
Sequence 47	1747	1747	0	0	0
Sequence 48	1747	1747	0	0	0
Sequence 49	1747	1747	0	0	0
Sequence 50	1747	1747	0	0	0
Sequence 51	1747	1747	0	0	0
Sequence 52	1747	1747	0	0	0
Sequence 53	1747	1747	0	0	0
Sequence 54	1747	1747	0	0	0
Sequence 55	1747	1747	0	0	0
Sequence 56	1747	1747	0	0	0
Sequence 57	1747	1747	0	0	0
Sequence 58	1747	1747	0	0	0
Sequence 59	1747	1747	0	0	0
Sequence 60	1747	1747	0	0	0
Sequence 61	1747	1747	0	0	0
Sequence 62	1747	1747	0	0	0
Sequence 63	1747	1747	0	0	0
Sequence 64	1747	1747	0	0	0
Sequence 65	1747	1747	0	0	0
Sequence 66	1747	1747	0	0	0
Sequence 67	1747	1747	0	0	0
Sequence 68	1747	1747	0	0	0
Sequence 69	1747	1747	0	0	0
Sequence 70	1747	1747	0		

D	b		1	gctcttggaactatttctcctctctctattcttccttagtcaaaagtaatttgttgattt	60
Q	y		1	GCYCTTTGGGATCATTTTCCCTCTCTCTATTTCCTTAGTCCGTAAAAGTTAAITTTGTGATTT	60
D	b		61	tgcgagccatgggtgtctcatcgcgcggtttctctgctgtcaaaagctgcgtttctcgacgc	120
Q	y		61	TTCGGAGCCATGGTGTTCTATCGGGGTTTTCTCTGCTGTCAAAAGCTGCGTCTCTCGAGC	120
D	b		121	gtccaacacagaaaaatttagcaacctbtgtgcggttggtttcaagtccaaaacctctcttggt	180
Q	y		121	GTCACACAGACAAATCTTAGCAACTCTGTGCGGTGGCTTCAAGTCCAACACCTCTCTGGT	180
D	b		181	cttgatctcgcttctgagctgcagaatgatgccagaaacacaggatcgccataaagaag	240
Q	y		181	CTTGATCTGGCTTCTCAGCTGCCAGANNATTGATTCAGAACAAACAGGATCGCCTTAAGAAG	240
D	b		241	ctcaagtcagagcatgaaaggttcaaattgggaaacatcacagttgatgtgtcttggt	300
Q	y		241	CTCAAAGTCAGAGCATGGAAGGTTCAAITGGGAAAAATCACAGTTGATATGGTCTCTGGT	300
D	b		301	qqaataagaggaatacagagattactgtggaacaccttacttaccoccatcaaqaa	360
Q	y		301	GGAAATCAGAGGAATACAGATTACTGTGGGAACCTCATTTACGCCGATGAAGGA	360
D	b		361	attcgcttctggggttgtctatcatgaatgccaaaaggtattacctgcagcaagcct	420
Q	y		361	ATTTCGCTTCGGGGCTTGCTATCTATGAATSCCAAAAGGTATTACCTGCAGCAAGACCT	420
D	b		421	gggggagagcccttgctgaaggtctctctggtctctttaacaggaagaagtgccaaca	480
Q	y		421	GGGGGAGAGCCCTTGCTGAAAGTCTCTCTGGCTTCCTTTAAACAGGAAGGTGCCATCA	480
D	b		481	aaagagcaagtggtatcatagtctcaggaattgcgaagtcgtactgtcccgcatacat	540
Q	y		481	AAAGAGCAAGTGGATTCAITGTCTCAGGAATTCGGAAGTCGTGCTACTGTCCCCGATCAT	540
D	b		541	gtatacaaaactattaatgccttacagtcacagctcatcaatgactcagtttgcctact	600
Q	y		541	GTATACAAACTATTGATGCCTTACCAGTCACAGCTCATCCAATGACTCAGTTTGTCTACT	600
D	b		601	ggagtcatggctcttcaggttcaaaagtgaattcaaaagggcatatgagaagaaggtatcac	660
Q	y		601	GGAGTCATGGCTCTTCAGGTTCAAAGTGAAITTCAAAGGCATATGAGAAAGGATTCAC	660
D	b		661	aatatcaagttatggaaacccacatatagaggattccatbgagtttgattgctcaagttccca	720
Q	y		661	AAATCAAAGTTATGGGAACCGCATATGAGGATTCATGAGTTTGAITGTCTCAAGTTCCA	720
D	b		721	cttggtgctgttatgttcatcgaggaatgtacaagaagccaacatatacctaagaagat	780
Q	y		721	CTTGTGTCTGCTTATGTTTATCGCAGGATGTACAAGAACGGCAACACTATPACCTAAGGAT	780
D	b		781	gactcactggattatgggtcgaattttgctcacatgcttggttttcagtagctctgacatg	840
Q	y		781	GACTCACTGGATTATGGTGCAAATTTTGCTCACATGCTTGGTTTCAGTAGCTCTGACATG	840
D	b		841	catgaagcttataaagctctatgtcacgatacacagtgatcatgaagtggttaacgctcaet	900
Q	y		841	CATGAGCTTATGAAGCTCTATGTACGATTACACAGTGAATCATGAAGGTGGTAACGTCAGT	900
D	b		901	gctcacacaggtcaacttggtgctagtgctttgtcagacccttacctctccttcgctgct	960
Q	y		901	GCTCACACAGGTCCACTTGGTTGCTAGTGTGTTTGTCTAGACCCCTTACCTCTCCTTCGCTGCT	960
D	b		961	gctttgaatggtttaactggacaccttcatggttttagccaatcaggaagattttcctatgg	1020
Q	y		961	GCYTTTGAATGGTTTATGCTGGACCACTTCATGTTTGTAGCCAATCAGGAAGTTTTCCTATGG	1020
D	b		1021	atcaaatctgtgtlagaggagtggtggggagacaatttccaagaagcaggttcaagaagctac	1080

Qy	1021	ATCAAAATCTGTTAGAGGAGTGTGGGGAGAACATTTCCAAAGAGCAGCTTGCAAGACATAC	1080
Db	1081	gcttgaaacaactgaaagtggaaggttgctccctgggtttcgacacatggaggttctgcgc	1140
Qy	1081	GCTTGGAAAACATTGAAGAGTGGCAAGGTTGTCCCTGTTTCGGACATGGAGTTCTCGCG	1140
Db	1141	aaagactaatccaaagatacacatgccaaagagaaggttcgctttgaagcattttccctgaagat	1200
Qy	1141	AAGACTGATCCAGATACACATGCCAGAGAGAGTTGCGTTTGAAGCATTTGCCCTGAAGAT	1200
Db	1201	ccactggttcaactgggttgcaaaactctacgaaggtgtccctccaatctctacagaacttg	1260
Qy	1201	CCACTGTTTCACTGGTGTCAAAACTCTACGAAGTGTTCCTCCAATCTTTCACAGAACTTG	1260
Db	1261	gcaagttaaaccccttgccaaatgttgatgccacagtggtgtgtgtgttaaacattat	1320
Qy	1261	GCAAGTTAAACCCTTGGCCAAATGTTGATGCCACAGTGGTGTGTGTTGTCGAACATTAT	1320
Db	1321	ggtttaactgaagcaagatatatacgggtccctcttgggtgatcaagagctcttggcatt	1380
Qy	1321	GGTTTAACTGAACAGAGATATTATACGGTCTCTTTTGGTGTATCAAGAGCTCTTGGCATT	1380
Db	1381	tgctctcagctaatthyggaccgagcctcttggatggccactagagagccaaagagtgct	1440
Qy	1381	TGCTCTCAGCTAATTTGGGACCGAGCTCTTGATTTGCCACTAGAGAGGCCAAAGAGTCT	1440
Db	1441	acaatggagtggtctgagaaaccattgcaagaagcatgattgtttgaaatctctgcgag	1500
Qy	1441	ACAAATGGAGTGGCTTGAGAACCACTTGCAAGAAGCATGATTTGTTTGAATCTCTCGGAG	1500
Db	1501	cataaagcacaatgtaaaactcttatgaataatttgcttgagaagcagtttttcttg	1560
Qy	1501	CATAAAAGCAACATGTAAATCTTTATGAATTAATTGCTTGAGAAAGCAGTTTTTCTTGG	1560
Db	1561	agccaaggtagctcgcattagatattcatcgatggcttagtcacgatttgaagattt	1620
Qy	1561	AGCAAGGTAGTGGCATTAGGATGTTATCGATTTGGCTTAGTACGGTTTTTGAAGAATTT	1620
Db	1621	tggttgtgtattttcagtttcggtttttaaagtataaccaatccttatcgatataaa	1680
Qy	1621	TGGTTGTGTAATTTTCAGTTTCGGTTTAAATAATGTTATACCAATACCTTATTCGATATAA	1680
Db	1681	tccaataattcgatttttacttttcttgtaaaaaaaacaaaaaataaaaaa	1740
Qy	1681	TTCAATATGATTCGATTTTTTACTTTTGTGTGAAAAAAACAAAAAATAAAAAA	1740
Db	1741	aaaaaaa 1747	
Qy	1741	AAAAAAA 1747	

RESULT

RESULT 2
ID T04199 standard: cDNA to mRNA. 1891 BP

ID T04199 standard; cDNA to hRNA; 1891 BP.
AC T04199:

AC 104199;
DT 25-JAN-1996 (first entry)

DE Potato citrate synthase cDNA.

KW Citrate synthase;

OS *Solanum tuberosum*.

FH	Key	Location/
----	-----	-----------

FT cds

ET
/*tag=

PN WO9524487-A.

PD 14-SEP-1995.

PF 07-MAR-1995; E00859.

PR 09-MAR-1994; DE-408629.

PR 22-SEP-1994; DE-435366.

PR 19-OCT-1994; DE-438821.
DA /ACRE \ HOGHOUSE-COVERING ACREAGE CMBW

PA (AGRE) HOECHS
DT Ia Cognat an

PI La Cognata U, Lan
PI Mueller-roeber B:

PI MUELLER-ROEBER
DR WPT: 95-328278

DR P-PSDB; R82838.
PT DNA encoding plant citrate synthase - used to regulate flower formation

PR 05-SEP-1997; US-057584.
 PR 05-SEP-1997; US-057629.
 PR 05-SEP-1997; US-057642.
 PR 05-SEP-1997; US-057645.
 PR 05-SEP-1997; US-057648.
 PR 05-SEP-1997; US-057651.
 PR 05-SEP-1997; US-057662.
 PR 05-SEP-1997; US-057668.
 PR 05-SEP-1997; US-057762.
 PR 05-SEP-1997; US-057765.
 PR 05-SEP-1997; US-057771.
 PR 05-SEP-1997; US-057776.
 PR 06-JUN-1997; US-048876.
 PR 06-JUN-1997; US-048880.
 PR 06-JUN-1997; US-048883.
 PR 06-JUN-1997; US-048892.
 PR 06-JUN-1997; US-048895.
 PR 06-JUN-1997; US-048898.
 PR 06-JUN-1997; US-048901.
 PR 06-JUN-1997; US-048917.
 PR 06-JUN-1997; US-048963.
 PR 06-JUN-1997; US-048971.
 PR 06-JUN-1997; US-049019.
 PR 06-JUN-1997; US-049374.
 PR 05-SEP-1997; US-057627.
 PR 05-SEP-1997; US-057634.
 PR 05-SEP-1997; US-057643.
 PR 05-SEP-1997; US-057646.
 PR 05-SEP-1997; US-057649.
 PR 05-SEP-1997; US-057654.
 PR 05-SEP-1997; US-057666.
 PR 05-SEP-1997; US-057760.
 PR 05-SEP-1997; US-057763.
 PR 05-SEP-1997; US-057769.
 PR 05-SEP-1997; US-057774.
 PR 05-SEP-1997; US-057777.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
 PI Fan P, Feng P, Ferric AM, Fischer CL, Florence C,
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 DR WPI: 99-053965/05.
 DR P-PSDB: W88711, W89037, W89038, W89039, W89040.
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS Claim 4: Page 440-441: 772pp: English.
 CC The invention relates to nucleic acid sequences (W84411 to W84633)
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC
 CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also diagnose:
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischaemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
 CC or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents a gene encoding a human secreted protein
 CC (see descriptor line for gene number and clone identification).

SQ Sequence 1848 BP; 575 A; 345 C; 476 G; 452 T;
 Query Match 2.5%; Score 43; DB 56; Length 1848;
 Best Local Similarity 97.8%; Pred. No. 1.44e-06;
 Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1799 cttttgtttgaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1843
 QY 1703 CTTTGTGTTGAAAAAAAAACAAAAAAAAAAAAAAAAAAAA 1747
 RESULT 13
 ID V54124 standard; DNA; 2519 BP.
 AC V54124:
 DT 17-DEC-1998 (first entry)
 DE Human membrane protein BA0306 coding sequence.
 KW Membrane protein; BA0306; BA2303; arteriosclerosis; coronary restenosis;
 KW therapy; human; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..1785
 FT /tag= a
 FT /transl_except= (pos: 289..291, aa: Xaa)
 FT /transl_except= (pos: 1015..1017, aa: Xaa)
 FT /transl_except= (pos: 1492..1494, aa: Xaa)
 FT /transl_except= (pos: 1540..1542, aa: Xaa)
 FT /transl_except= (pos: 1582..1584, aa: Xaa)
 FT /transl_except= (pos: 1609..161, aa: Xaa)
 FT /transl_except= (pos: 1693..1695, aa: Xaa)
 FT /note= "Xaa= unspecified amino acid"
 PN WO9838305-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1998; J00835.
 PR 25-FEB-1998; JP-062263.
 PR 28-FEB-1997; JP-062259.
 PR (NISB) JAPAN TOBACCO INC.
 PI Nakamura Y, Tanaka T, Tsukada S;
 DR WPI: 98-481206/41.
 DR P-PSDB: W74580.
 PT Membrane protein(s) BA0306 and BA2303 - useful for, e.g. treatment
 PT and prevention of arteriosclerosis and restenosis
 PS Claim 16: Page 90-95; 141pp; Japanese.
 CC This sequence encodes the human BA0306 membrane protein of the invention.
 CC The invention also relates to the human BA2303 membrane protein. The two
 CC membrane proteins are specifically expressed in mammals during
 CC arteriosclerosis and coronary restenosis. The membrane proteins,
 CC fragments of them, and antibodies against them are useful in the
 CC treatment and prevention of arteriosclerosis and restenosis. Transgenic
 CC mice expressing the extracellular region of the membrane proteins are
 CC useful as models for studying these disorders.
 CC Sequence 2519 BP; 766 A; 439 C; 492 G; 809 T;
 SQ Sequence 2519 BP; 766 A; 439 C; 492 G; 809 T;
 Query Match 2.5%; Score 44; DB 50; Length 2519;
 Best Local Similarity 92.3%; Pred. No. 5.39e-07;
 Matches 48; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 2419 tttttactttgtttgaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2470
 QY 1696 TTTTACTTTGTGTTGAAAAAAAAACAAAAAAAAAAAAAAAAAAAA 1747
 RESULT 14
 ID V84530 standard; DNA; 472 BP.
 AC V84530:
 DT 01-MAR-1999 (first entry)
 DE Human secreted protein gene 120 clone HPD20.
 KW Human secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyrold; digestion;

W P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 22:08:38 1999; MasPar time 126.78 Seconds
1191.820 Million cell updates/sec
Tabular output not generated.

Title: >US-08-702-718-5
Description: (1-1747) from US08702718.seq
Perfect Score: 1747
N.A. Sequence: 1 GCTCTTGGGATCTATTTCCT.....AAAAA.....TTTTTTTTTTTTTT
Comp: CGAGACCCCTAGATAAGGA.....TTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued

Statistics: Mean 9.068; Variance 5.849; scale 1.550

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	48	2.7	7218	2	US-08-232-Sequence 14, Applicati	7.31e-13
2	47	2.7	7218	2	US-08-232-Sequence 14, Applicati	2.63e-12
3	40	2.3	700	1	US-07-845-Sequence 1, Applicati	1.72e-08
4	40	2.3	700	1	US-08-469-Sequence 1, Applicati	1.72e-08
5	40	2.3	1046	3	US-08-361-Sequence 4, Applicati	1.72e-08
6	40	2.3	1046	3	US-08-484-Sequence 4, Applicati	1.72e-08
7	40	2.3	1080	1	US-08-225-Sequence 1, Applicati	1.72e-08
8	41	2.3	3933	4	PCT-US95-0 Sequence 1, Applicati	5.00e-09
9	41	2.3	3933	4	US-08-199-Sequence 1, Applicati	5.00e-09
10	40	2.3	7859	1	US-07-854-Sequence 4, Applicati	1.72e-08
11	40	2.3	7859	3	US-08-450-Sequence 15, Applicati	1.72e-08
12	38	2.2	84	3	US-08-738-Sequence 3, Applicati	1.98e-07
13	38	2.2	84	3	US-08-664-Sequence 97, Applicati	1.98e-07
14	38	2.2	85	1	US-08-120-Sequence 97, Applicati	1.98e-07
15	38	2.2	85	3	US-08-478-Sequence 97, Applicati	1.98e-07
16	38	2.2	350	1	US-08-171-Sequence 14, Applicati	1.98e-07
17	39	2.2	836	3	US-08-698-Sequence 7, Applicati	5.85e-08
18	39	2.2	880	3	US-08-616-Sequence 7, Applicati	5.85e-08
19	39	2.2	1454	3	US-08-713-Sequence 7, Applicati	5.85e-08
20	38	2.2	1978	5	Patent No. 5188642-3	1.98e-07

21 38 2.2 5852 1 US-07-867-Sequence 2, Applicatio 1.98e-07
22 36 2.1 120 3 US-08-151-Sequence 28, Applicati 2.20e-06
23 36 2.1 120 1 US-08-153-Sequence 28, Applicati 2.20e-06
24 36 2.1 144 2 US-08-702-Sequence 26, Applicati 6.63e-07
25 37 2.1 208 2 US-08-686-Sequence 37, Applicati 6.63e-07
26 37 2.1 216 2 US-08-686-Sequence 34, Applicati 6.63e-07
27 36 2.1 222 4 PCT-US93-0 Sequence 15, Applicatio 2.20e-06
28 37 2.1 347 1 US-08-104-Sequence 2, Applicatio 6.63e-07
29 36 2.1 893 3 US-08-798-Sequence 66, Applicati 2.20e-06
30 36 2.1 1013 1 US-08-314-Sequence 30, Applicati 2.20e-06
31 36 2.1 1013 1 US-07-920-Sequence 30, Applicati 2.20e-06
32 36 2.1 1013 1 US-08-086-Sequence 23, Applicati 2.20e-06
33 37 2.1 3073 4 PCT-US91-0 Sequence 30, Applicati 6.63e-07
34 36 2.1 3073 1 US-07-688-Sequence 31, Applicati 6.63e-07
35 36 2.1 3200 1 US-08-384-Sequence 1, Applicatio 2.20e-06
36 36 2.1 3200 1 US-08-444-Sequence 1, Applicatio 2.20e-06
37 37 2.1 3214 1 US-08-484-Sequence 17, Applicati 6.63e-07
38 37 2.1 3214 1 US-08-484-Sequence 17, Applicati 6.63e-07
39 36 2.1 3581 3 US-08-738-Sequence 1, Applicatio 2.20e-06
40 36 2.1 4104 2 US-08-453-Sequence 94, Applicati 2.20e-06
41 36 2.1 4104 4 PCT-US93-1 Sequence 94, Applicati 6.63e-07
42 37 2.1 6671 3 US-08-555-Sequence 1, Applicatio 6.63e-07
43 37 2.1 6671 4 PCT-US95-0 Sequence 1, Applicatio 6.63e-07
44 37 2.1 6671 1 US-08-280-Sequence 1, Applicatio 6.63e-07
45 37 2.1 6671 2 US-08-457-Sequence 1, Applicatio 6.63e-07

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC: 7218 BP.
AC xxxxxx
DE Sequence 14, Application US/08232463
CC Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: 30472/114 INMU
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 INMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: pTZqpt-Fls
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 2.7%; Score 48; DB 2: Length 7218;
Best Local Similarity 3.6%; Pred. No. 7.3le-13;
Matches 7, Conservative 113; Mismatches 72; Indels 0; Gaps 0;

Db 1059 CTGGGATATTCCTCTCTCTATTCCTCCCTAGTAGTAAAGTAAATTTGTTG 53
QY 4 CTGGGATATTCCTCTCTCTATTCCTCCCTAGTAGTAAAGTAAATTTGTTG 53
Db 1119 YY 1178
QY 64 CGAGCCATGCTGTCTATCGCGCGGTTCTCTCTGTCAAAGCTGGTTCTCGAGCGGTC 123
Db 1179 YY 1238
QY 124 CAACAGACAATCTTAGCAACTCTGTGCGGTGCTTCAAGTCCAACTCTTCTGCTCT 183
Db 1239 YYYYYYYYYYYY 1250
QY 184 GATCTGGTTCT 195

RESULT 2
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx

DE Sequence 14, Application US/08232463
CC Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232,463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935,313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 25-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: pTZqpt-Fls
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 2.7%; Score 47; DB 2: Length 7218;
Best Local Similarity 4.3%; Pred. No. 2.63e-12;
Matches 13; Conservative 160; Mismatches 126; Indels 0; Gaps 0;

Db 1156 YY 1215
Cc 494 TCCACTTGCTCTTTGATGGCACCTTTCCTGTTAAAGAGCCAGAGACCTTCAGGC 435
Db 1216 YY 1275
Cc 434 AAGGCTCTCCCGAGGCTTTGCTGCAGGTAATACCTTTTGGCATTATAGATAGACAAG 375
Db 1276 YY 1335
Cc 374 CCGCGAAAGCAATTCCTTCATCGGGTCAAGTAATGAGGTTTCCACAGTAATCCTGTC 315
Db 1336 YY 1395
Cc 314 ATTCTCTCATTCACCAAGAACCATATCAACTGTGATGTTTCCCAATTGACCTTTCCA 255
Db 1396 YY 1454
Cc 254 TGCTCTGACTTGAGCTCTTTTAGCGGATCCTGTTGTTCTGCAATCAATCTTCGAGCTC 196

RESULT 3
ID US-07-846-992-1 STANDARD; DNA; UNC; 700 BP.
AC xxxxxx

DE Sequence 1, Application US/07846992
CC Sequence 1, Application US/07846992
CC Patent No. 5583046
CC GENERAL INFORMATION:
CC APPLICANT: Valenta, Rudolf
CC APPLICANT: Duchene, Michael
CC APPLICANT: Pettenburger, Karin
CC APPLICANT: Breitenbach, Michael
CC APPLICANT: Kraft, Dietrich
CC APPLICANT: Rumpold, Helmut
CC APPLICANT: Schelner, Otto
CC TITLE OF INVENTION: Birch Pollen Allergen P14 for Diagnosis
CC TITLE OF INVENTION: and Therapy of Allergic Diseases
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/846,992
CC FILING DATE: 19920606
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/353,844
CC FILING DATE: 18-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jones III, Harry C
CC REGISTRATION NUMBER: 20,280

CC REFERENCE/DOCKET NUMBER: 6530-011
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-8864/9741
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 700 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORGANISM: Betula verrucosa
CC IMMEDIATE SOURCE:
CC SEQUENCE 700 BP; 212 A; 127 C; 181 G; 180 T; 0 OTHER.
SQ
Query Match 2.3%; Score 40; DB 1; Length 700;
Best Local Similarity 97.6%; Pred. No. 1.72e-08;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 654 TTGTTGAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAA 695
|||||
QY 1706 TTGTTGAAAAA...ACAAAAA...AAAAAAAAAAAA 1747
|||||
RESULT 4
ID US-08-469-555-1 STANDARD; DNA; UNC; 700 BP.
AC xxxxxx
DT
DE Sequence 1, Application US/08469555
CC Sequence 1, Application US/08469555
CC Patent No. 5648242
CC GENERAL INFORMATION:
CC APPLICANT: Valenta, Rudolf
CC APPLICANT: Duchene, Michael
CC APPLICANT: Pottenburger, Karin
CC APPLICANT: Breitenbach, Michael
CC APPLICANT: Kraft, Dietrich
CC APPLICANT: Rumpold, Helmut
CC APPLICANT: Scheiner, Otto
CC TITLE OF INVENTION: Birch Pollen Allergen p14 for Diagnosis
CC TITLE OF INVENTION: and Therapy of Allergic Diseases
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/469,555
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/846,992
CC FILING DATE: 06-JUN-1992
CC APPLICATION NUMBER: US/07/353,844
CC FILING DATE: 18-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jones III, Harry C
CC REGISTRATION NUMBER: 20,280
CC REFERENCE/DOCKET NUMBER: 6530-011
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090

CC TELEFAX: (212) 869-8864/9741
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 700 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORGANISM: Betula verrucosa
CC IMMEDIATE SOURCE:
CC LIBRARY: POLLEN FROM ALLERSON AB, ENGELHOLM, SWEDEN
CC SEQUENCE 700 BP; 212 A; 127 C; 181 G; 180 T; 0 OTHER.
SQ
Query Match 2.3%; Score 40; DB 1; Length 700;
Best Local Similarity 97.6%; Pred. No. 1.72e-08;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 654 TTGTTGAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAA 695
|||||
QY 1706 TTGTTGAAAAA...ACAAAAA...AAAAAAAAAAAA 1747
|||||
RESULT 5
ID US-08-361-467B-4 STANDARD; DNA; UNC; 1046 BP.
AC xxxxxx
DT
DE Sequence 4, Application US/08361467B
CC Sequence 4, Application US/08361467B
CC Patent No. 5633441
CC GENERAL INFORMATION:
CC APPLICANT: De Greef, Willy
CC APPLICANT: Van Emmelo, John
CC APPLICANT: De Oliveria, Dulce E.
CC APPLICANT: De Souza, Maria-Helena
CC APPLICANT: Van Montagu, Marc
CC TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR
CC TITLE OF INVENTION: EMBRYOS
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
CC STREET: P.O. Box 1404
CC CITY: Alexandria
CC STATE: Virginia
CC COUNTRY: United States
CC ZIP: 22313-1404
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/361,467B
CC FILING DATE: 22-DEC-1994
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/681,492
CC FILING DATE: 04-APR-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/EP90/01275
CC FILING DATE: 01-AUG-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 89 402 224.3
CC FILING DATE: 04-AUG-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Schulman, Robert M.
CC REGISTRATION NUMBER: 31,196
CC REFERENCE/DOCKET NUMBER: 010830-027
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 836-6620

CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/738,367
CC FILING DATE:
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brown, Scott A.
CC REGISTRATION NUMBER: 32,724
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 498-8224
CC TELEFAX: (617) 876-5851
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna
CC SEQUENCE 84 BP; 76 A; 0 C; 0 G; 7 T; 1 OTHER.
SQ
Query Match 2.2%; Score 38; DB 3; Length 84;
Best Local Similarity 93.0%; Pred. No. 1.98e-07;
Matches 40; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 1 TTTTITNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43
QY 1705 TTITGTTGAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAA 1747
RESULT 13
ID US-08-664-596B-3 STANDARD; DNA; UNC; 84 BP.
AC xxxxxx
DT
DE Sequence 3, Application US/08664596B
CC Sequence 3, Application US/08664596B
CC Patent No. 5807703
CC GENERAL INFORMATION:
CC APPLICANT: Jacobs, Kenneth
CC APPLICANT: McCoy, John
CC APPLICANT: Lavallie, Edward
CC APPLICANT: Racie, Lisa
CC APPLICANT: Merberg, David
CC APPLICANT: Treacy, Maurice
CC APPLICANT: Evans, Cheryl
CC APPLICANT: Spaulding, Vikki
CC APPLICANT: Bowman, Michael
CC TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
CC TITLE OF INVENTION: ENCODING THEM
CC NUMBER OF SEQUENCES: 37
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc.
CC STREET: 87 CambridgePark Drive
CC CITY: Cambridge
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02140
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/664,596B
CC FILING DATE:
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brown, Scott A.
CC REGISTRATION NUMBER: 32,724
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 876-5851
CC TELEFAX: (617) 876-5851
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna
CC SEQUENCE 84 BP; 76 A; 0 C; 0 G; 7 T; 1 OTHER.
SQ

CC LENGTH: 84 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna
CC SEQUENCE 84 BP; 76 A; 0 C; 0 G; 7 T; 1 OTHER.
SQ
Query Match 2.2%; Score 38; DB 3; Length 84;
Best Local Similarity 93.0%; Pred. No. 1.98e-07;
Matches 40; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 1 TTTTITNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43
QY 1705 TTGTTTGAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAA 1747
RESULT 14
ID US-08-120-827-97 STANDARD; DNA; UNC; 85 BP.
AC xxxxxx
DT
DE Sequence 97, Application US/08120827
CC Sequence 97, Application US/08120827
CC Patent No. 5525495
CC GENERAL INFORMATION:
CC APPLICANT: KEENE, JACK D.
CC APPLICANT: KING, PETER H.
CC APPLICANT: LEVINE, TODD
CC TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
CC TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
CC TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
CC NUMBER OF SEQUENCES: 101
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC ADDRESSEE: P.C.
CC STREET: 1755 Jefferson Davis Highway, Fourth Floor
CC CITY: Arlington
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22202
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/120,827
CC FILING DATE: 15-SEP-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Oblon, No. 5525495man F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 714-158-0 CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)413-3000
CC TELEFAX: (703)413-2220
CC TELEEX: 248855 OPAT UR
CC INFORMATION FOR SEQ ID NO: 97:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 85 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: cdna
CC SEQUENCE 85 BP; 50 A; 5 C; 10 G; 20 T; 0 OTHER.
SQ
Query Match 2.2%; Score 38; DB 1; Length 85;
Best Local Similarity 88.0%; Pred. No. 1.98e-07;
Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 35 TTTAAATTTTGTATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 84
QY 1698 TTTTACTTTTGTGAAAAAAAAACAAAAAAAAAAAAAAAAAAAA 1747

W P S R L A (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Oct 23 21:11:35 1999; MasPar time 2958.33 Seconds
Tabular output not generated. 1383.723 Million cell updates/sec

Title: >US-08-702-718-5
Description: (1-1747) from US08702718.seq
Perfect Score: 1747
N.A. Sequence: 1 GCTCTTGGATCTATTTCCT.....AAAAA.....AAAAA 1747
Comp: CGAGAACCTAGATAAGGA.....TTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est11
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 12.404; Variance 6.933; scale 1.789

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	247	14.1	471	20	AA896953 L30-386T3 Ice plant	9.20e-150
2	210	12.0	502	17	T44185 7448 Lambda-PRL2 Arabi	1.67e-122
3	190	10.9	417	24	C96653 Rice callus Ory	6.97e-108
4	187	10.7	478	31	R90544 16899 Lambda-PRL2 Arab	1.07e-105
5	187	10.7	620	26	AU002699 AU002699 Bombyx mori p	1.07e-105
6	174	10.0	676	23	A1176862 EST220465 Normalized r	2.94e-96
7	172	9.8	527	31	R90561 16916 Lambda-PRL2 Arab	8.24e-95
8	163	9.3	585	15	AI05406966 vm94h09.r1 Knowles Sol	2.60e-88
9	155	8.9	582	22	AI054878 coau002314 Cotton Bol	1.48e-82
10	152	8.7	307	10	AA231705 CDO534.R cDNA from oat	2.10e-80

LOCUS	1	AA896953	471 bp	mRNA	EST	06-APR-1998
DEFINITION	L30-386T3 Ice plant Lambda Uni-zap XR expression library, 30 hours NACI treatment Mesembryanthemum crystallinum cDNA clone L30-386 5', similar to Citrate synthase, mitochondrial precursor, mRNA					
ACCESSION	AA896953					
NID	AA896953					
VERSION	AA896953.1	GI:3033346				
KEYWORDS	EST.					
SOURCE	common ice plant.					
ORGANISM	Mesembryanthemum crystallinum					
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: euphyllophytes; Spermatophyta: Magnoliophyta: eudicotyledons; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.					
AUTHORS	Cushman, J.C.					
TITLE	An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum					
JOURNAL	Unpublished (1997)					
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2150267.					

ALIGNMENTS

11	143	8-2	522	10	AA237580	mx11d01.r1 Soares mous	5-69e-74
12	134	7-7	501	11	AA313713	EST185581 Colon carcin	1-42e-67
13	131	7-5	213	8	Z17455	ATTS0029 AC16H Arabido	1-89e-65
14	122	7-0	540	24	A1238915	GH15054 5Prime GH Dros	4-14e-59
15	121	6-9	473	14	AA498291	VH38g11.r1 Barstead mo	2-08e-58
16	115	6-6	252	17	AA720224	33417 Lambda-PRL2 Arab	3-27e-54
17	114	6-5	466	36	AA112941	zn59b06.r1 Stratagene	1-63e-53
18	112	6-4	444	10	AA245100	mx08g10.r1 Soares mous	4-01e-52
19	110	6-3	418	34	W40867	mc66e02.r1 Soares mous	9-83e-51
20	110	6-3	503	9	AA178724	mt16h01.r1 Soares mous	9-83e-51
21	110	6-3	509	9	AA177425	mt33d11.r1 Soares mous	9-83e-51
22	108	6-2	704	9	AA192035	rs04h04.r1 Sommer Pris	2-39e-49
23	107	6-1	397	14	C28571	C28571 Rice callus cDN	1-18e-48
24	102	5-8	489	23	A1168709	ox56a01.sl Soares NHM	3-29e-45
25	102	5-8	494	18	AA830955	ox56a01.sl NCI-CCAP GC	3-29e-45
26	101	5-8	510	21	C88687	C88687 Mouse early dia	1-60e-44
27	101	5-8	566	15	AA568724	nm21f09.sl Soares mous	1-60e-44
28	100	5-7	455	13	AA427173	ves3c01.r1 Soares mous	7-77e-44
29	100	5-7	506	26	AA901227	UI-R-A1-dp-g-06-0-UI.s	3-76e-43
30	99	5-7	752	25	A127328	mp70f10.x1 Soares 2NB	8-77e-42
31	97	5-6	331	11	AA332541	ESN36505 Embryo, 8 wee	8-77e-42
32	96	5-5	321	19	F11594	HSC22A101 normalized i	4-22e-41
33	96	5-5	343	21	AA992542	ot88g03.sl Soares tota	4-22e-41
34	96	5-5	442	33	W01297	yz94h09.r1 Soares mela	4-22e-41
35	96	5-5	452	9	AA196933	zq59h08.r1 Stratagene	4-22e-41
36	94	5-4	425	33	W14146	mb19e03.r1 Soares mous	9-70e-40
37	95	5-4	642	13	C23820	C23820 Dictyostelium d	2-02e-40
38	94	5-4	672	23	A1176797	EST220393 Normalized r	9-70e-40
39	92	5-3	338	11	AA337465	EST42223 Endometrial t	2-21e-38
40	93	5-3	481	9	AA172950	ms20b08.r1 Stratagene	4-63e-39
41	91	5-2	530	13	AA416759	zu09a03.sl Soares_test	1-05e-37
42	90	5-2	611	26	AU000363	AU000363 Bombyx mori p	4-98e-37
43	89	5-1	360	35	C13265	C13265 Yuji Kohara unp	2-35e-36
44	89	5-1	592	32	N21557	YX60b08.sl Soares mela	2-35e-36
45	88	5-0	461	9	AA196934	zq59h09.r1 Stratagene	1-11e-35

Contact: Cushman JC

Department of Biochemistry and Molecular Biology

Oklahoma State University

350 Noble Research Center, Stillwater, OK 74078-3035, USA

Tel: 405-744-6207

Fax: 405-744-7799

Email: jcushman@biochem.okstate.edu

PCR Primers

FORWARD: T7

BACKWARD: T3

Plate: L30-4 row: H column: 6

Seq primer: T3

High quality sequence stop: 320.

FEATURES

Location/Qualifiers

1..471
/organism="Mesembryanthemum crystallinum"
/note="Vector: Lambda Uni-zap XR, Bluescript SK; Site_1:
EcoRI; Site_2: XhoI"
/db_xref="taxon:3544"
/clone="L30-386"
/clone_lib="Ice plant Lambda Uni-zap XR expression
library, 30 hours NaCl treatment"
/tissue_type="Leaf, 30 h 0.4M NaCl"
/dev_stage="Six week old"
127 a 95 c 122 g 127 t

BASE COUNT

127 a 95 c 122 g 127 t

ORIGIN

Query Match 14.18; Score 247; DB 20; Length 471;
Best Local Similarity 80.3%; Pred. No. 9.20e-150;
Matches 354; Conservative 0; Mismatches 83; Indels 4; Gaps 4;
Db 3 CATTGTGGGTAGCGGACATTCAGATCCCTACCTTTCATTTGCAGCTGCACCTGAATGGG 62
QY 913 CACTTGGTCTAGTGTCTGACACCCCTACCTCTCCTCGCTGCTTGAATGT 972
Db 63 TTGCTGGCCCACTTATGATGATGAGCAATCAGGAAGTGTGTGTGGATCAAGTCAGTT 122
QY 973 TTAGCTGGACCACTTCATGTGTTTGGCAATCAGGAAGTGTGTGTATGGATCAAACTCTGT 1032
Db 123 GTGATCAATGTGGAGAAACATCACACGAGCAGCTAAAGATTTATCTCGGAGACA 182
QY 1033 GTAGAGAGTGTGGGAGAACATTTCCAAAGAGCAGTGTGAAGACTACGCTTGGAAAAA 1092
Db 183 TTAAGAGTGGCAAGGTGTACCTGGATATGGCATGGAGTTCGCGCAAGCGGATCCA 242
QY 1093 TTGAAGAGTGGCAAGGTGTCCCTGGTTTCGGACATGGAGTTCGCGCAAGCTGATCCA 1152
Db 243 AGATACCTTGGCCAGAGGACTTGGCCCTTAAGCACTTACCTAATGATCCACTATTTCAA 302
QY 1153 AGATACACATGCCAGAGAGAGTTCGCTTTGAAGCAATTTGCTCAAGATCCACTGTTTCAA 1212
Db 303 CTGTTTTCGAAGCTGTATGAATGGTGCTCCAAATTCGACAGAGCTTGGGAAGTAAAGA 362
QY 1213 CTGTTTGCAGAACTCTACGAAGTGTTCCTCCAAATTCACAGAACTTGGCAAGTTAA-A 1271
Db 363 ACCATGGCCGAATTTGATGCCACAGTGGGTGTATTTGCTGAACACTACTATGTTTGACAA 422
QY 1272 CCCTTGGCCAAATTTGATGCCACAGTGG-TGTGTTGTTGAACACTATTATGTTTAACT- 1329
Db 423 GAAGCAAGATACACTGT 443
QY 1330 GAA-GCAAGATATTATACGGT 1349

RESULT 2
LOCUS 744185 502 bp mRNA EST 07-JAN-1998
DEFINITION 7448 Lambda-PRL2 Arabidopsis thaliana cDNA clone 123C9T7, mRNA
sequence.
ACCESSION T44185
NID g2758988
VERSION T44185.1 GI:2758988
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 502)
AUTHORS Newman, T., deBruin, F. J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
Retzel, E., and Somerville, C.
TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729

COMMENT

On Jan 7, 1998 this sequence version replaced gi:948518.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: 17.

FEATURES

Location/Qualifiers

1..502

source

/organism="Arabidopsis thaliana"
/strain="var. Columbia"
/note="Vector: lambda zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliquae). The vector is BRL's lambda zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.
/db_xref="taxon:3702"
/clone="123C9T7"

BASE COUNT 125 a 90 c 118 g 149 t 20 others
ORIGIN

Query Match 12.0%; Score 210; DB 17; Length 502;
Best Local Similarity 79.1%; Pred. No. 1.67e-122;
Matches 284; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Db 8 TGGATTATGGTGCAAAATTTTCCACATGTTGGGATTTGATGATGAAAGGTGAAAGAGC 67
QY 788 TGGATTATGGTGCAAAATTTTCCACATGTTGGTGTGTTTTCAGTAGCTCTGCATGATGAGC 847
Db 68 TCATGAGGCTTTACATACCATCCACAGTATCATGAAGTGGAAATGTTAGTGCTCACA 127
QY 848 TTATGAAGCTCTATGTACAGATACACAGTATCATGAAGTGGTAACTGATGCTCACA 907
Db 128 CTGTCACCTGGTTGGTAGTCACACTTTTACAGCCCATATCTGTCAATTTGACAGCTGCATTAA 187
QY 908 CAGGTCACCTTGGTTGCTAGTGTGTTGTGTCAGACCCTTACCTCTCTCTGCTGCTGCTTGA 967
Db 188 ATGTTTACCTGGGCCACTCCATGTTGGCTAATCAGGAAGTTTTCCTTTGGATCAAT 247
QY 968 ATGTTTACCTGGGCCACTCCATGTTGGCTAATCAGGAAGTTTTCCTTTGGATCAAT 1027
Db 248 CAGTCGTAGAGGAATGTGGAGAAGATATATCAAAAGAACAGCTTGAAGAATATATGTTGGG 307
QY 1028 CTGTTGTAGAGAGCTGTGGGAGAACATTTCCAAAGAGCAGTTGAAGACTACCTTGA 1087
Db 308 AAACATTAAACAGTGGCAAGTTTATTCGGGGTATGTCACGGTGTCTTCCCAATACT 366
QY 1088 AAACATTGAAAGTGGCAAGTTTTCCTGCTGTTCCGACATGGAGTTCTGCGCAAGACT 1146

RESULT 3
LOCUS C96653 417 bp mRNA EST 19-OCT-1998
DEFINITION C96653 Rice callus Oryza sativa cDNA clone C10531_5A, mRNA
sequence.
ACCESSION C96653
NID g3758954
VERSION C96653.1 GI:3758954
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.


```

REFERENCE 1 (bases 1 to 417)
AUTHORS   Sasaki,T. and Minobe,Y.
TITLE     Rice cDNA from callus
JOURNAL   Unpublished (1994)
COMMENT   On Aug 21, 1998 this sequence version replaced.

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = "RGP".

FEATURES             Location/Qualifiers
     source
       1..417
         /organism="Oryza sativa"
         /strain="cultivar Nipponbare, sub-species Japonica"
         /note="Vector: pBluescript II SK+; Site.1: SalI; Site.2:
         NotI; cDNA prepared from rice callus mRNAs by using
         oligo(dT) as a primer and ligating to the SalI-NotI site
         of pBluescript II SK+ phagemid."
         /db_xref="taxon:4530"
         /map="11"
         /clone="C10531_5A"
         /clone_lib="Rice callus"
BASE COUNT  115 a  76 c  108 g  118 t
ORIGIN
Query Match      10.9%; Score 190; DB 24; Length 417;
Best Local Similarity 73.4%; Pred. No. 6.97e-108;
Matches 306; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

Db 2  TAGTTTACCGGAGGATCTTCAAGGATGGGAAACTATAGACGCTGATATGCATCGAC 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 733 TAGTTTATTCGAGGATGTACAGAAGCGCAACACTATACCTTAAGATGACCTGGAT 792
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62  TAGTCAGCAAACTTTTCACACATGCTTGGGTTTGATGATGCCAAATGCTCGAGTTGAT 121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 793 TAGTGTGCAAAATTTTCTGCATGCTTTGGTTTCAGTAGCTCTGACATGATGCTTAIG 852
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 CGACTATGTGACATGCACACTGATCATGATGAGGTGGAAATGTGAGTCTCATCTGGA 181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 853 AAGCTCTATGTCACGATACACAGTATCATGAAGTGTGTAAGTGCAGTCTCACACAGT 912
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 CATCTGTTGGAGAGTCTCTGTGACATCCTTATCTCTTTTTCAGCTGCATGATGTT 241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 913 CACTTGTGCTAGTGTCTGTGACACCTTACCTCTCTCTGCTGCTTGAATGTT 972
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 TTAGCTGGACCAATTCACGCGCTGCTTAATCAGGAAGTCTGTTTGGATCAAAATCTGTA 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 TTAGCTGGACCACTTCATGTTTAGCAATCAGGAAGTTTTCATGATGATCAAAATCTGTT 1032
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 ATAGTGTGACAGGATGATGATGTACAACTGATCAACTCAAAAGATGATGTGGAAGACA 361
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1033 GTAGAGGAGTGTGGGGAGAACATTTTCAAGAGAGCAGTTGAAAGACTACGCTTGGAAACA 1092
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 CTAAAGTGGAAAGTGTCTCTGCT -CGGTCATGAGGTTCTACGTAAGACCGAT 417
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1093 TTGAAAGTGGCAAGTGTCTCTGCTTTTCGGACATGGAGTCTCTCGGCNAGACTGAT 1149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
LOCUS      R90544
DEFINITION 16899 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189D1577, mRNA
sequence.
ACCESSION  R90544
NID        9958084
VERSION    R90544.1
KEYWORDS   EST.
SOURCE     thale cress.

```

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 478)
AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
Retzel,E. and Somerville,C.
TITLE     Genes galore: a summary of methods for accessing results from
JOURNAL   large-scale partial sequencing of anonymous Arabidopsis cDNA clones
MEDLINE   Plant Physiol. 106, 1241-1255 (1994)
COMMENT   95148729
          On Apr 14, 1993 this sequence version replaced gi:693655.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313ton@bm.cl.msu.edu
Seq primer: T7 dye primer.
          Location/Qualifiers
FEATURES             1..478
     source
       /organism="Arabidopsis thaliana"
       /strain="var columbia"
       /note="Vector: lambda Zip-Lox; Site.1: Sal; Site.2: Not;
       Lambda PRL2 is a cDNA library derived from equal
       quantities of 4 pools of mRNA. The mRNA sources were 1) 7
       day germinated etiolated seedlings; 2) tissue culture
       grown roots; 3) staged plants half with 24 hour light
       cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
       same plants as 3 but aerial tissue (stems, flowers and
       siliques. The vector is BRL's lambda Zip-Lox. The cDNA
       inserts were directionally cloned with Sal-Not arms using
       oligo dt primed cDNA.
       /db_xref="taxon:3702"
       /clone="189D157"
       /clone_lib="Lambda-PRL2"
BASE COUNT  119 a  90 c  113 g  137 t  19 others
ORIGIN
Query Match      10.7%; Score 187; DB 31; Length 478;
Best Local Similarity 79.3%; Pred. No. 1.07e-105;
Matches 265; Conservative 0; Mismatches 67; Indels 2; Gaps 2;

Db 4  TCAGACCCATATCTGTCATTTCAGCTGCATTAATGGTTTACGTGGGCCACTCCATGGT 63
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 934 TCAGACCCCTTACCTCTCTTCGCTGCTTGAATGGTTTACGTGGGCCACTTCATGTT 993
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64  TTGGCTAATCAGGAAGTTTTCGTTTGGATCAAAATCAGTCGTAGAGGAATGTGGAGAAGAT 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 994 TTAGCCAAATCAGGAAGTTTTCGTCATGATCAATCTGTTGTAGAGAGTGTGGGAGAAC 1053
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 ATATCAAAAGAACAGTTGAAAGATATGTTTGGAAAACATTAACACAGTGGCAAGTTATT 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1054 ATTTCCAAAGACGAGTTGAAAGACTACGCTTGGAAAACATTTGAAAGTGGCAAGTTGTC 1113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 CCGGGATATGGTCAGGTTTCTGCGCAATCTGATCCAAAGATATGATGCCAAGAGAA 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1114 CCGTGGTTGCGACATGAGGTTCTGCGCAAGACTGATCCAAAGATACATGCCAGAGAG 1173
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 TTTGCCCTGAAGCATCTACCTGACGNCCTCTTTTTCAGCTGGTGTCAAAGCTTTATGAA 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1174 TTCGCTTTGAAGCATTTGCCCTGAAGATCCACTGTTTCAACTGGTTGCCAAACTCTAGCA 1233
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 GTTGTGCTCCCTGTTCTCACTGAGCTTTGGAAA 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1234 GT-GTTCCTCAA-TTCTTACAGAAGTGGCAA 1265
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5

```



```

Db 483 GTCGTAACTCTCATCTGACACGCTTTTCCGACTTCCTTCTGTAGTGGTCAGCCA 542
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1077 GCTTTCACTGCTCTTTGGAATGTTCTCCCACTCCTCTACAAGATTTGATCCA 1018
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GACAAGCACTCTGTTTGTAGTCCATGACAGAGCCCTGCCAGCCCATTCATGCTGC 602
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1017 TAGCAAAACTTCTGATTTGGCTAAAGCATGAAGTGGTCAGCTAAACCATTCAAAAGCAGC 958
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 TGCANAGGACAGTAAGGCTCTGAAGGCGCTGCCCAAGTGGCTGTGTGGGCACT 662
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 957 AGCGAAGAGAGTAAGGCTCTGACAAAGCACTAGCAACCAAGTGACCTGTGTGAGCACT 898
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 TACATTGCCACC 674
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 897 GACGTTACCACC 886
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
LOCUS 16916 Lambda-PRL2 Arabidopsis thaliana cDNA clone 189H3T7, mRNA
DEFINITION sequence.
ACCESSION R90561
NID R958101
VERSION R90561.1 GI:958101
KEYWORDS EST.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 527)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
Retzel,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT On Apr 14, 1993 this sequence version replaced gi:693674.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.
FEATURES
source
1. .527
/organism="Arabidopsis thaliana"
/strain="var columbia"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.
/db_xref="taxon:3702"
/map="873F09: 1q21.3-1q23.2; 9"
/clone="189H3T7"
/clone_lib="Lambda-PRL2"
BASE COUNT 130 a 106 c 122 g 148 t 21 others
ORIGIN
Query Match 9.8% Score 172; DB 31; Length 527;
Best Local Similarity 76.7% Pred. No. 8,24e-95;
Matches 247; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```

```

Db 8 TGATGCGCAGAAAGTATTACCTACTGCCCCAGTCTGGAGCTGAACCAATTACCGAGGTCT 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 387 TGAATGCCAAAGAGTATTACCTGCGCAAGAGCTTGGGGAGAGCCCTTGGCTGAAGGTCT 446
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 TTTTGGCTTCTCTTAAGTGAAGGTACCTAGCAAGAGCAAGTTGAAGCACTGTGCA 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 447 TCTGTGCTTCTTTAACAGGAAAGTGCCATCAAAAGAGCAAGTGGATTCTTGTCTCA 506
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 128 AGATTGGCGAACCGTCTGCTGTGCGCAGATTATGTACAAATGCCATCATCTCTGCC 187
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 507 GGAATTGCGAAGTGTGCTACTGTCCCGCATGTATATACAAACTATTATGATCCCTTACC 566
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 TTCACAGCTCATCAATGACTCAATTTGCTAGCGGTGTTATGGCCCTCCAGGTGCAAG 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 567 AGTCACAGCTCATCAATGACTCAATTTGCTAGCGGTGTTATGGCCCTCCAGGTGCAAG 626
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 TGATTTCAAAGGCATATGAGATGGAATTCATTAAGTTCAGAGTTCGGGAGGCAACATA 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 627 TGAATTTCAAAGGCATATGAGAAAGGATTCACAAATCAAGTTATGGAACGACATA 686
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 TGAGGATTGCTCAACCTGATT 329
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 687 TGAGGATTCATGAGTTGATT 708
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS AA606966 657 bp mRNA EST 30-SEP-1997
DEFINITION vm94109.r1 Knowles Solter mouse blastocyst B1 Mus musculus CDNA
clone IMAGE:1005953 5' similar to SW:ClSY_PIG P00889 CITRATE
SYNTASE, MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION AA606966
NID G2455859
VERSION AA606966.1 GI:2455859
KEYWORDS EST.
SOURCE house musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 657)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1328460.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:570169
High quality sequence stop: 473.
FEATURES
source
1. .657
/organism="Mus musculus"
/strain="B6D2 F1/J"
/note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGTGCACCTGCACCGTTTCTTTTCTTTT-3' cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (life
technologies). Two different size selections: B1 (larger
inserts) and B3."
/db_xref="taxon:10090"
/clone="IMAGE:1005953"

```

```

/clone_lib="Knowles Solter mouse blastocyst B1"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
BASE COUNT      161 a 167 c 181 g 148 t
ORIGIN

```

```

Query Match      9.3%; Score 163; DB 15; Length 657;
Best Local Similarity 66.6%; Pred. No. 2,60e-88;
Matches 397; Conservative 0; Mismatches 192; Indels 7; Gaps 6;

Db 19 GAGCTCATGCGTTTACCTCACCATCATAGTACCATGAGGGTGTAAATGTAAGTGCC 78
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 844 GAGCTTATGAAGCTCTATGTCACGATACACAGTATCATGAAGGTGTAAGCTCAGTGCT 903
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 CACACAAGCCATTGTTGGGAGAGCCATT-TCAGACCCTTACCTGTCTCTTTCAGACAGCC 137
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 904 CACAGAGTCACTGGTGGTCTAGTCTTTGTCAGACCCTTACCTCTCGCTCGCTGCT 963
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 ATGAATGGGCTGGGGGCT-CTACATGGACTAGCAAAATCAGGAGGTGCTTGTCTGGCTG 196
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 964 TTGAATGGTTAGCTGGACCCTTCTATGGTTTAGCCATCAGGAAGTTTCTTAAGATC 1023
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 ACACAGCTACAGAGAAAGTTGGCAAGACGTTGTCAGATGAGAAATTACGAGATACATC 256
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1024 AAATCTGTTGTAGAGAGTGTGGGAGAACATTTCCAAAGAGCAGCTTGAAGAGCTACGCT 1083
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 TGAACACACTCAATTCAGGAGGGTGTCCAGGATACGCTATGCTAGTACTCAGGAAG 316
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1084 TGAACAACTTGAAGAGTGGCAAGTTTGTCCCTGGTTTCGGACATGGAGTTCTGCGCAAG 1143
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 ACTGACCTCGCTATTCTGTGAGGAGAGATTGTCTGAAACATCTGCCTAAGGGATTC 376
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1144 ACTGATCCAAGATACATGCGCAGAGAGATTGCTTTGAAGCATTTGCCTGAAG-ATCC 1202
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 CCAATGTTCAAGCTGGTGTGCTGAGTGTACAGATTGTGCCCAATATCCTCTTACAGCAAG 436
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1203 ACTGTTTCAA-CTGGTTGCAAACTCTACGA-AGTGTTCCTCCCAATCTTACAGAACTG 1260
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 437 GGAAGCTTAAGAACCTTGGGCCCAAGCTAGAGCTCACAGTGGGTGCTGCCAGTACT 496
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1261 GCAAGATTAA--ACCTTGGCCAAATGTTGATGCCACAGTGGTGTGTTGTAACATT 1318
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 ATGGATCAGCGAGATGAATCTACTACAGTCTCTGTTTGGAGTGTCTCGGCACTGGGTG 556
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1319 ATGGTTTAACTCAAGCAAGATATTATACGGTCTCTTTGGTGTATCAAGAGCTCTTGCA 1378
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 TGTAGCCAGCTCATCTGGAGCAGCCCTAGGCTTCCCTGGAAAGGCCCAAG 612
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1379 TTGTCTCAGCTAATTTGGGACCGAGCTCTTGGATTGCCACTAGAGAGG-CAAG 1434
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
LOCUS AI054878 585 bp mRNA EST 16-JUL-1998
DEFINITION coau0002G14 Cotton Boll Abscission Zone cDNA Library Gossypium
            hirsutum cDNA clone coau0002G14 5', mRNA sequence.
ACCESSION AI054878
NID 93325992
VERSION AI054878.1 GI:3325992
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          euryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
          Rosidae; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 585)
AUTHORS Wan, C.-H., Yu, Y., Sasinowski, M., and Wing, R. A.
TITLE Cotton EST Database: Sequence Analysis of 2000 cDNA Clones from an
        Abscission Zone Library
JOURNAL Unpublished (1998)
COMMENT On Jan 9, 1998 this sequence version replaced gi:930638.
        Contact: Wing RA

```

```

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: SP030 (AACAGCTATGACCATGATTA)
High quality sequence stop: 239.
Location/Qualifiers
1..585
/organism="Gossypium hirsutum"
/notes="Vector: Uni-ZAP XR; Site 1: EcoRI; Site 2: XhoI.
This is a Uni-ZAP XR custom cDNA library made by
Stratagene (U.S.A.: 1-800-424-5444); Stratagene cat.
#837201."
/db_xref="taxon:3635"
/clone="coau0002G14"
/lab_host="XLI-Blue MRF"

```

FEATURES

```

Source
1..585

```

```

BASE COUNT      144 a 120 c 141 g 172 t      8 others
ORIGIN

```

```

Query Match      8.9%; Score 155; DB 22; Length 585;
Best Local Similarity 73.7%; Pred. No. 1.48e-82;
Matches 291; Conservative 0; Mismatches 97; Indels 7; Gaps 7;

```

```

Db 50 GCATCGTGTGTTTGAAGGAGTATCAGCGCTATCCAGTCAAGTCACTGTTAAACAG 109
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 67 GCATCGTGTGTTTATCCGCGGTTTCTGCTGTCAAGCTCGGTTCTCGAGGGTCCAA 126
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 CATTCCAGTCTCAGCAATCTGTTAGTGGCTTCAATGCACTGCTCTTCCGATTTGAT 169
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 127 CACACAAATCTTAGCAACTCTGTGCGTGTCAAGTCCAAACCTCTTCTGCTTGAT 186
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 CTTATTCGCGAGTTGAAGAAATTGATTCCTGAACACAGAGCGGCTGAAGAACTGAGG 229
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 187 CTCGCTTGTGAGTGTGAAGAAATTGATTCAGAACACAGAGATCGCTTAAAGAGCTCAAG 246
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 CGAAGCTGNAAGAACTGCAACTTGGGAATATCTGTTGATATGCTATGTTGGAATG 289
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 247 TCAGAGCATGGAAGGTTCAATTTGGGAACATCAGATTTGATATGTTCTTGGTGAATG 306
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 AGAGGATGACAGNGTTGCTTTGGGAAACCTCATTTACTTG-CCCAGATGGGGGAAATCG 348
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 307 AGAGGAATGACAG-GATTACTGTGGGAAACCTCATTTACTTGACCCGATGAAGGAATCG 365
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 CTTTGGGCTTCTCATTCGCG-AGTGTCAAAACCTATTACCAGCTGCAAA-CCTCATCG 406
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 356 CTTTCGGGCTTGTCTATCTATGAATGCCAAAGGTATTACCTGCAAGAACCTGGGGG 425
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 AGAACCTT-CGTNAGGCTTT-TATGGCTTCTTT 438
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 426 AGAGCCCTTGCCTGAAGGCTTCTCTGCTTCTTT 460
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
LOCUS AA231705 307 bp mRNA EST 27-FEB-1997
DEFINITION CDO534.R cDNA from oat Avena sativa cDNA clone CDO534, mRNA
            sequence.
ACCESSION AA231705
NID g1854091
VERSION AA231705.1 GI:1854091
KEYWORDS EST.
SOURCE oat.
ORGANISM Avena sativa
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          euryliophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
          Poaceae; Avena.
REFERENCE 1 (bases 1 to 307)
AUTHORS Vanbeyne, A.E., Sorrells, M.E., Park, W.D., Ayres, N.M., Fu, H.,
        Cartinhour, S.W., and McCouch, S.R.
TITLE Anchor Probes for Comparative Mapping of Grass Genera
JOURNAL Unpublished (1997)

```


vector: Lambda ZAPII; Physiological condition: cycling cells."

Best Local Similarity 83.4%; Pred. No. 1.89e-65;
Matches 176; Conservative 0; Mismatches 33; Indels 2; Gaps 2;
BASE COUNT 59 a 40 c 52 g 62 t
ORIGIN

Query Match 7.5%; Score 131; DB 8; Length 213;

Best Local Similarity 83.4%; Pred. No. 1.89e-65;
Matches 176; Conservative 0; Mismatches 33; Indels 2; Gaps 2;

Db 5 TCACATCCACAGTATGATGAGGTGGAAATTTAGTGTCTACACTGTGTACCTGGTTG 64
QY 863 TCACGATACAGTATGATGAGGTGGAAATTTAGTGTCTACACTGTGTACCTGGTTG 922

Db 65 GTAGTCACATTCAGACCCATATCTCATTTGCACCTGCATTAATAGTTTGTAGCTGGC 124
QY 923 CTAGTGTCTTGTGAGACCCCTTACCTCTCCTTGCCTGCTGCTTTGAAATGTTTGTAGCTGGAC 982

Db 125 CACTCCATGTTTGGTGAATCAGGAAGTTTGGTGGATCAA-TCAGTGTGAGGAAAT 183
QY 983 CACTTCATGTTTGGTGAATCAGGAAGTTTGGTGGATCAA-TCAGTGTGAGGAAAT 1042

Db 184 GTGAGAGATATATC-AAAGAACAGTTGAA 213
QY 1043 GTGGGAGACATTTCCAAAGAGCAGTTGAA 1073

RESULT 14
LOCUS AI238915 540 bp mRNA EST 18-NOV-1998
DEFINITION GH15054.5prime GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH15054 5prime, mRNA sequence.

ACCESSION AI238915
NID 3833773
VERSION AI238915.1 GI:3833773
KEYWORDS EST.
SOURCE fruit fly.

ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 540)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
On Aug 21, 1998 this sequence version replaced.

CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estfruitfly.berkeley.edu
Plate: 150 row: E column: 6
High quality sequence stop: 419.
Location/Qualifiers
1. .540

FEATURES
source
/organism="Drosophila melanogaster"
/note="Organ: head; Vector: pot2; Site:1: EcoRI; Site:2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."
/db_xref="taxon:7227"
/clone="GH15054"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
BASE COUNT 103 a 183 c 156 g 98 t
ORIGIN

Query Match 7.0%; Score 122; DB 24; Length 540;

Best Local Similarity 63.7%; Pred. No. 4.14e-59;
Matches 283; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Db 97 CTGATTTGGTCGGCAACTTTGTGAAGATGCTGGCTACGACACAGCCGCCCTTCACCGAG 156
QY 787 CTGGATTATGTCGCAAAATTTTGTCTACATGCTTGTTCAGTAGCTGTGACATGATGAG 846

Db 157 CTGATGCTCTCTATCTGACCATCCACAGTGACACAGGGTGGCAAGCTGTCTGCCAC 216
QY 847 CTTATGAAGCTCTATGTCACGATACACAGTGATCATGAAGTGGTGAAGCTGCTGCTAC 906

Db 217 ACCGTTCACTTGGTGGCTCCGCCCTCAGGATCCCTTACCTCTCTTTTGGCGCGGCTG 276
QY 907 ACAGTCACTTGGTGGCTAGTGTCTTGTTCAGACCCCTTACCTCTCTTCTGCTGCTTGG 966

Db 277 AACGGTCTGGCTGCTCCCTTCACGGCTGCCCAACAGGAGGTGCTGCTGTGCTGCTGCC 336
QY 967 AATGTTAGTGTGACCACTTTCATGGTTAGCCCAATCAGGAAGTTTGTCTATGGATCAA 1026

Db 337 AAGTGTGAGAGAGGGCGGCAACACCGTCCGAGGAGCAGCTCAAGGAGTACATCTGG 396
QY 1027 TCTGTTAGAGGAGTGTGGGAGAACATTTCCAAAGCAGATTGAAGACTAGCTTGG 1086

Db 397 AAGACCTTCAAGTCCGACAGGTGTTCCGGCTACCGACACGCCGCTCTCCGCAAGACC 456
QY 1087 AAAACATTGAAAAGTGGCAAGTTTCCCTGTTTCGGACATGGAGTTCTGCGCAAGACT 1146

Db 457 GATCCCGCTACCTGCCACGCTGAGTTCGCGTGAAGCACCCTGCCGAGGAGGACG 516
QY 1147 GATCCAGATACATCCAGAGAGTTCGCTTTGAAGCATTTGCTGAAGATCCACTG 1206

Db 517 TTCAGCTGGTGTGGAAGATCTAC 540
QY 1207 TTTCACCTGGTGTGCAAACTCTAC 1230

RESULT 15
LOCUS AA498291 473 bp mRNA EST 01-JUL-1997
DEFINITION vh38g11.r1 Barstead mouse pooled organs MPLR84 Mus musculus CNA clone IMAGE:889316 5' similar to SW:CI5Y.PIG P00889 CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR ; mRNA sequence.

ACCESSION AA498291
NID 92233314
VERSION AA498291.1 GI:2233314
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 473)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402331.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:517276
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 472.
Location/Qualifiers
1. .473

FEATURES
source

```
/organism="Mus musculus"
/strain="FVB/N"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGAATCTGAGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[GTTCATTCGGTACC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."
/db_xref="taxon:10090"
/clone="IMAGE:889316"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
```

```
BASE COUNT      120 a      122 c      124 g      107 t
ORIGIN

Query Match      6.9%; Score 121; DB 14; Length 473;
Best Local Similarity 68.2%; Pred. No. 2.08e-58;
Matches 264; Conservative 0; Mismatches 119; Indels 4; Gaps 4;

Db 38 GAGCTCATGGTTTGTTACCTCACCATGATGACCATGAGGTTGTAATGTAAGTGC- 96
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 844 GAGCTTATGAAGCTCTATGTCAGGATACACAGTATGTAAGGTTGTAAGTGCAGTGC 903
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 97 CACACAAGCCATTGGTGGGACGCGCCTT-TCAGACCCCTTACCTGTCCTTTGCAGCAGCC 155
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 904 CACACAGTCACTTGGTTGCTAGTGTGTTGTCAGACCCCTTACCTCTCCTCGCTGCT 963
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 156 ATGAATGGCTGGCGGGCT-CTACATGGACTAGCAATCAGGAGGTGCTTGTCTGGCTG 214
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 964 TTGAATGGTTTACCTGGACCACTTCATGGTTTAGCCAATCAGGAAGTTTGTCTATG 1023
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 215 ACACAGCTACAGAAGAAAGTTGGCAAGACGTGTCAGATGAGAAGTTACGAGACTACATC 274
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1024 AAATCTGTTGTAGAGAGTGTGGGAGAACATTTCCAAGAGCAGTTGAAAGACTACGCT 1083
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 275 TGAACACACTCAATTCAGAGCGGTGGTCCA-GGATAGGTCATGCAGTACTAGGAGAG 333
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1084 TGAACAACTTCAAAAGTGGCAAGTTGTCCCTGGTTTCGGACATGGAGTTCTGCGCAAG 1143
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 334 ACTGACCTCGCTATTCTCTGTCAGGAGAGTTTGTCTGTAACATCTGCTAAAGATCCC 393
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1144 ACTGATCCAAGATACACATGCCAGAGAGAGTTCCGTTTGAAGCAITTCCTGAAGATCCA 1203
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 394 ATGTTCAAGCTGGTGGCTCAGCTGTAC 420
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1204 CTGTTTCAACTGGTTGCAAAACTCTAC 1230
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: Sat Oct 23 22:01:18 1999
Job time : 2983 secs.

M P E R F H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 16:59:48 1999; MasPar time 18.52 Seconds
Tabular output not generated. 538.628 Million cell updates/sec

Title: >US-08-702-718-6
Description: (1-469) from US08702718.pep
Perfect Score: 3443
Sequence: 1 MVFYRGVSLLSKLPSPAVQQPLERPKSVTMEWLENHCKKA 469

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 35.933; Variance 168.408; scale 0.213

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	3443	100.0	469 14	R82840 Tobacco citrate synth
2	3098	90.0	471 15	R86383 Potato citrate synth
3	3098	90.0	471 14	R82838 Potato citrate synth
4	2800	81.3	437 14	R82839 Sugar beet citrate sy
5	313	9.1	436 3	R14357 Citrate synthase enco
6	116	3.4	456 34	W3570 Methyl tetrahydropter
7	106	3.1	220 9	R47340 Fragment of chromaffi
8	103	3.0	329 29	W55330 H. pylori ORF hp3e110
9	103	3.0	329 24	W55574 H. pylori ORF 06ep302
10	102	3.0	355 24	W29179 Rat CC chemokine rece
11	101	2.9	3084 35	W50891 Mouse laminin A chain
12	98	2.8	355 11	R52749 C-C chemokine recepto
13	98	2.8	355 24	W25751 Human MIP-1alpha/RAN
14	98	2.8	355 25	W26588 Human MIP-1 alpha/RAN
15	95	2.8	2164 2	P81045 Sequence of the viral
16	95	2.8	2164 1	P80131 Peptides translated f

Peptide encoded by pl 1.04e-02
PSQM525-derived pep 1.04e-02
M. vaccae potD homolo 1.04e-02
Arabidopsis chloropia 8.95e-01
Klebsiella pneumoniae 1.20e-02
Acetobacter xylinum b 1.04e-02
O8 ethylene response 8.95e-01
A. thaliana ethylene 8.95e-01
Human placenta G-prot 1.39e-02
T-lymphocyte IL15A An 1.85e-02
Human T-lymphocyte TL 1.85e-02
Nocardia corallina re 1.39e-02
Nocardia corallina al 1.85e-02
Prostaglandin-EP3-1 1.85e-02
Prostaglandin-EP3-21 1.85e-02
Prostaglandin-EP3-9 r 1.85e-02
Human EP3-V receptor. 1.85e-02
Salmonella secreted p 1.39e-02
H. pylori secreted or 1.60e-02
H. pylori secreted or 1.60e-02
Mature Penicillin V a 1.39e-02
Soluble type I insuli 1.85e-02
Full length Penicilli 1.39e-02
Human CLOCK protein. 1.60e-02
Sequence of beta-chai 1.85e-02
CF-5 pathogen resista 1.85e-02
IGF-I receptor. 1.85e-02
Rat type I insulin-li 1.39e-02

ALIGNMENTS

RESULT 1
ID R82840 standard; Protein; 469 AA.
AC R82840:
DT 25-JAN-1996 (first entry)
DE Tobacco citrate synthase.
KW Citrate synthase; flower formation.
OS Nicotiana tabacum
PN W09524487-A.
PD 14-SEP-1995.
PF 07-MAR-1995; E00859.
PR 09-MAR-1994; DE-408629.
PR 22-SEP-1994; DE-435366.
PR 19-OCT-1994; DE-438821.
PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
PI La Cognata U, Landschuetze V, Muller-rober B, Landschuetze V;
PI Mueller-roeber B;
DR WPI; 95-328278/42.
DR N-PSDB; T04201.
PT DNA encoding plant citrate synthase - used to regulate flower formation,
PT to improve storage of tubers, etc. and to reduce sprouting
PS Disclosure; Page 60-63; 87pp; English.
CC To identify a cDNA from tobacco which codes for citrate
CC synthase, a cDNA bank of leaf tissue from tobacco was prep'd.
CC Plaque of this cDNA bank were screened using a radioactive DNA
CC probe which comprises Solanum tuberosum citrate synthase cDNA
CC (T04199). One of the clones was sequenced. The nt. sequence is
CC given in T04201.
SQ Sequence 469 AA;

Query Match 100.0%; Score 3443; DB 14; Length 469;
Best Local Similarity 100.0%; Pred. No. 7.69e+299;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mvfyrqvsllsklsravqqtlnslsvrwlvqvqtssgldlrslqelipeqqdrkkls 60
Qy 1 MVFYRGVSLLSKLSRAVQQTNLNSVRFWLVQVQTSSGLDRSELQELIPEQQDRKKLS 60
Db 61 ehgkvqlgnitvdmvlgmrgmtgllwetslldpdegirfgrlsieyckvlpakpgge 120
Qy 61 EHGKVLQGNITVDMVLGMRGNTGLLWETSLLDPEDEIRFGRLSIYECQKVLPAKPGGE 120

Db 121 plpeqllwllltqkpskeqvdselsqelsratvpdhyvktidaltptahpmtqfatgvm 180
 QY 121 PLPEGLLWLLLTGKVPSEKQVDSLSQELRSRATVPDHYVKTIDALPVTAHPTQFATGVM 180
 Db 181 alqvsefokayekgihksklyeptyedsmsliagvplvaayvyrmyknqntipkddsl 240
 QY 181 ALOVQSEFOKAYEKG IHKSKLWEPTYEDSMSLIAQVPLVAAYVYRMYKNQNTIPKDDSL 240
 Db 241 dyganfahmlgfsdssdmhlmklyvtihdsheggvnsahtghlvasalsdpylsfaaain 300
 QY 241 DYGANFAHMLGFSDDSMHLMKLYVTIHSDEGGVNSAHTGHLVASALSDPYLSFAAALN 300
 Db 301 giagphlqanqevllwiksvveecgeniskeqlkdyawktlkskvpgfghvrlrkt 360
 QY 301 GLAGPHLQANQEVLLWIKSVVEECGENISKEQLKDYAWKTLLKSGVPGFGHVRLRKT 360
 Db 361 prytcrefalkhlpdpqlfqlvaklyevflqnlaklnpwpnvdaahsgvllnyygit 420
 QY 361 PRYTQREFALKHLPDPQLFQLVAKLYEVFLQNLAKLNPNWPNVDAHSGVLLNYYGIT 420
 Db 421 earyyvtlfgvsralgicsqlidwralglplerpksvmtmewlenhckka 469
 QY 421 EARYYTVLFGVSRALGICSQIWDRLALGLPLERPCKSVTMEWLENHCKKA 469

RESULT 2
 ID R86383 standard: Protein; 471 AA.

AC R86383;
 DT 24-APR-1996 (first entry)
 DE Potato citrate synthase.
 KW Citrate synthase; inhibitor; increased storage capacity; potato;
 KW Antisense DNA.
 OS Solanum tuberosum.
 PN DE4408629-A1.
 PD 14-SEP-1995.
 PF 09-MAR-1994; 408629.
 PP 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-Roeber B, Landschuetze V;
 DR WPI; 95-321536/42.
 DR N-PSDB; T03410.
 DT Inhibiting citrate synthase (CS) activity in plants - to inhibit
 PT flower formation and improve storage capacity, e.g. in potatoes,
 PT also new CS sense and anti-sense DNA sequences
 PS Claim 5; Page 15-19; 35pp; German.
 CC The potato citrate synthase (CS) gene or cDNA sequence encoding CS
 CC (T03410) may be used to produce antisense CS sequences. CS DNA
 CC sequences are useful for altering CS activity in plants. Antisense
 CC CS sequences can be used to inhibit CS expression in plants and has
 CC the effect of inhibiting flower formation and by doing so improves
 CC the plant's storage capacity. This is partic. useful in crop plants
 CC of any kind but esp. useful in potatoes. In addition to altering CS
 CC activity the DNA sequences can also be used to identify similar
 CC transgenic plants with altered CS activity. This sequence
 CC represents the amino acid sequence of potato citrate synthase.
 SQ Sequence 471 AA;

Query Match 90.0%; Score 3098; DB 15; Length 471;
 Best Local Similarity 89.6%; Pred. No. 3.55e-267;
 Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4.

Db 1 mvfyrsvllsklrsvavqsnvsnvrlvqvtssgldlrslvqelipeqgdrkkik 60
 QY 1 MVFYRVSLLSKLRSAVQQTNLNSVRLVQVTSGLDLRSEL-QELIPEQGDRLKKIK 59
 Db 61 sdm-ksginivtdmvlgmrgmtglwphylpdegirfgrlsipeckvlpakpgg 119
 QY 60 SEHGKVLQNITVDVYLGMRCMTGLLWETSLDPEGIRFRGLSIYECQKVLPAKPGG 119

Db 120 eplpeqllwllltqkpskeqvsnivsgiaesgiisliimyttdaltptahpmtqfatg 179
 QY 120 EPLPEGLLWLLLTGKVPSEKQVDSLSQELRSRATVPDHYVKTIDALPVTAHPTQFATG 178
 Db 180 vmalqvsefokayekgihksklyeptyedsmsliagvplvaayvyrmyknqntipkde 239
 QY 179 VMALOVQSEFOKAYEKG IHKSKLWEPTYEDSMSLIAQVPLVAAYVYRMYKNQNTIPKDD 238
 Db 240 sldygafahmlgfsdssdmhlmrlyvtihdsheggvnsahtghlvasalsdpylsfaa 299
 QY 239 SLDYGANFAHMLGFSDDSMHLMKLYVTIHSDEGGVNSAHTGHLVASALSDPYLSFAA 297
 Db 300 alnqlagphlqanqevllwiksvveecgeniskeqlkdyawktlkskvpgfghvrlr 359
 QY 298 ALNQLAGPHLQANQEVLLWIKSVVEECGENISKEQLKDYAWKTLLKSGVPGFGHVRLR 357
 Db 360 ktvprytcqrefalkhlpdpqlfqlvaklyevflqnlaklnpwpnvdaahsgvllnyy 419
 QY 358 KTDPRYTQREFALKHLPDPQLFQLVAKLYEVFLQNLAKLNPNWPNVDAHSGVLLNY 417
 Db 420 gltearyyvtlfgvsralgicsqlidwralglplerpksvmtmewlenhckka 471
 QY 418 GLTEARYYTVLFGVSRALGICSQIWDRLALGLPLERPCKSVTMEWLENHCKKA 469

RESULT 3

ID R82838 standard: Protein; 471 AA.

AC R82838;
 DT 25-JAN-1996 (first entry)
 DE Potato citrate synthase.
 KW Citrate synthase; flower formation; tuber storage.
 OS Solanum tuberosum.
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 DR WPI; 95-328278/42.
 DR N-PSDB; T04199.
 DT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Disclosure; Page 53-56; 87pp; English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of
 CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. A cDNA library was prepd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 471 AA;

Query Match 90.0%; Score 3098; DB 14; Length 471;
 Best Local Similarity 89.6%; Pred. No. 3.55e-267;
 Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4;

Db 1 mvfyrsvllsklrsvavqsnvsnvrlvqvtssgldlrslvqelipeqgdrkkik 60
 QY 1 MVFYRVSLLSKLRSAVQQTNLNSVRLVQVTSGLDLRSEL-QELIPEQGDRLKKIK 59
 Db 61 sdm-ksginivtdmvlgmrgmtglwphylpdegirfgrlsipeckvlpakpgg 119
 QY 60 SEHGKVLQNITVDVYLGMRCMTGLLWETSLDPEGIRFRGLSIYECQKVLPAKPGG 119
 Db 120 eplpeqllwllltqkpskeqvsnivsgiaesgiisliimyttdaltptahpmtqfatg 179
 QY 120 EPLPEGLLWLLLTGKVPSEKQVDSLSQELRSRATVPDHYVKTIDALPVTAHPTQFATG 178

Db 180 vmalgvsefkaeygkghskwyetvedsmnlaiaqvpvaayvirmvngdtipkde 239
 QY 179 VMAOVSEFQKAYEGKHKSLEPTVEDSMLSIAQVPLVAAVYRMYKNGNTIPKDD 238
 Db 240 slidyganfahmlgfsenhellmlytihsdheggvnsahtghlvasalsdpylsfaa 299
 QY 239 SLDYGANFAHMLGFSSEDMHEL-MKLYTYTIHSDHEGGVNSAHTGHLVASALSDPYLSFAA 297
 Db 300 ainglagplhglancgvlwksvveecgenisseqldkdyvktlinsgkvvggfhgvlr 359
 QY 298 ALINGLAGPLHGLANCGVLWKSVEECGENISKEQLADYAWKTLKSGKVVPFGHGVLR 357
 Db 360 ktvpyrtqcrefamkhlbedplfqlvsklyevlflnqlaklkwpnvndahsgvlinyy 419
 QY 358 KTDPRYTCOREFALKHLPEDPLFQLVAKLYEVFLQFLQNLAKLPWPVNDHSGVLLNYY 417
 Db 420 gltearytytlvgvralgicsqlwdralgplrpkpsvntwlenqckka 471
 QY 418 GLTEARYTYTLVGVSRAIGCSQLWDRALGPLRPKPSVTMEWLENCKKA 469
 RESULT 4
 ID R82839 standard; Protein; 437 AA.
 AC R82839;
 DT 25-JAN-1996 (first entry)
 DE Sugar beet citrate synthase.
 KW Citrate synthase; flower formation.
 OS Beta vulgaris strain Zuchtlinie 5S 0026
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR N-PSDB; T04200.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Disclosure; Page 57-60; 87pp; English.
 CC To identify a cDNA from sugar beet which codes for citrate
 CC synthase, a cDNA bank of leaf tissue from sugar beet was prep'd.
 CC Plaques of this cDNA bank were screened using radioactive DNA
 CC probes which comprise a mixture of Solanum tuberosum citrate
 CC synthase cDNA (T04199) and Nicotian tabacum citrate synthase cDNA
 CC (see T04201). One of the clones was sequenced. The nt. sequence is
 CC given in T04200.
 SQ Sequence 437 AA;
 Query Match 81.3%; Score 2800; DB 14; Length 437;
 Best Local Similarity 85.7%; Pred. No. 7.44e-240;
 Matches 373; Conservative 34; Mismatches 27; Indels 1; Gaps 1;
 Db 1 snldlrselbelpeqqrllkikkefsgfqlgnindvmlgmrmtellwetslldp 60
 QY 35 SSGDLRSLOBLIFEQQDRLKXKSEHGKVLGNITVDMVLGMRGTMGLTWETSLDDP 94
 Db 61 eegirfgrfsiepcqklpaasagaepipeglillwllltgkvpksqevdalsadirkra 120
 QY 95 DEGIRFGRFSIEPCQKVLPAAPGGEPLPEGLLWLLLTGKVPKQVDSLSQELSPATV 154
 Db 121 pdhvyktidalpitahmtqfctgvmalqtrsefcgkayekghkskfwepayedclsia 180
 QY 155 PDHVKYKTIDALPVTAHPTQFATGVMALQVSEFQKAYEKGHKSLEPTVEDSMLSIA 214
 Db 181 qpvvaayvirmvngkvipiddslidvggnfnahlgidsqgmlelmrlyvtihdshegg 240
 QY 215 QPVLVAAVYRMYKNGNTIPKDDSLIDYGANFAHMLGFSSEDMHMLKLYTYIHSDEGG 274
 Db 241 nvsahgthlvgsplsdpylsfaaalinglagplhglancgvlwksvveecgenisteql 300

QY 275 NVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVEECGENISKEQL 334
 Db 301 kdywktlinsgkvvpfqlgvirktdprvtcqrrefalkhlppddpfqlvsklyevvppil 360
 QY 335 KDYAWKTLKSGKVVPFGHGVLRKTDPRYTCOREFALKHLPEDPLFQLVAKLYEVFLQEL 394
 Db 361 lelgkvknwpnvndahsgvllnhhygitearytytlvgvrsigicsqlwdralgplr 420
 QY 395 QNLAKL-NPWPVNDHSGVLLNYYGITEARYTYTLVGVSRAIGCSQLWDRALGPLR 453
 Db 421 pksvtmewlekfckr 435
 QY 454 PKSVTMEWLENHCKK 468
 RESULT 5
 ID R14357 standard; Protein; 436 AA.
 AC R14357;
 DT 23-JAN-1992 (first entry)
 DE Citrate synthase encoded by aara.
 KW Acetic acid resistance.
 OS Acetobacter sp.
 PN J03219878-A.
 PD 27-SEP-1991.
 PF 05-FEB-1990; 024395.
 PR 15-FEB-1989; JP-033776.
 PR 05-FEB-1990; JP-024395.
 PA (NAKA-) NAKANO SUMISE KK.
 DR WPI; 91-329112/45.
 DR N-PSDB; Q14367.
 PT Acetic acid resistant gene with in plasmid and transformed
 PT acetobacter - improves yield of acetic acid fermentation.
 PS Disclosure; Fig 5; 12pp; Japanese.
 CC The aara gene encodes the citrate synthase and is part of an
 CC acetic acid resistance operon comprising aara, aarb and aarc.
 CC The DNA can be used to prepare an acetic acid resistant strain of
 CC bacteria for use in acetic acid fermentation.
 CC See also R14358 and 59.
 SQ Sequence 436 AA;
 Query Match 9.1%; Score 313; DB 3; Length 436;
 Best Local Similarity 30.3%; Pred. No. 1.06e-15;
 Matches 99; Conservative 87; Mismatches 109; Indels 32; Gaps 28;
 Db 91 eevyillngelpnkagdydtfntlnhtllheqinffngfrdhpmailcgtvgals 150
 QY 124 EGLLWLLLTCKVPSKEQVDSLSQELSRATVPDHVYKTTIDALPVTAHPMTQFATGVMALQ 183
 Db 151 a-f-ydpandialp-an-rdla---amrliakiptiaawayk--ytqaeafiyprnd-in 200
 QY 184 VQSEFQKAYEKGHKSLEPTVEDSMLSIAQVPLVAAVYRMYKNGNT-I-PKDDSLD 241
 Db 201 yaenflsmfmmfarnsepykvnpvlararnrllilhadheq-naststvtlagstganpfac 259
 QY 242 YGANFAHML-G-FSSS-DMHE-LMK-LY-VTI-HSDHEGGVNSAHTGHLVASALSDPYLS 294
 Db 260 iaagialawpahgaganeavlk-mlarigk-kenip-afiq-v-kdknsgvklmfgfhr 314
 QY 295 FAALNGLAGPLHGLANQEVLLWIKSVVEECGENISKEQLKDYAWKTLKSGKVVPFGHG 354
 Db 315 vyknfdprakimqqtcheviteigikddpildilavelekials-ddyfvqrklyppndfy 373
 QY 355 VLKRTDPRYTC-ORE-FA-LKHLP--EDPLFQLVAKLYEVFLQFLQNLAKLPWPVNDH 409
 Db 374 sqilkamgipstmf-tvlfavarttg 399
 QY 410 SGVLLNYYGITEARYTYTLVGVSRAIG 436
 RESULT 6
 ID W38570 standard; Protein; 456 AA.
 AC W38570;
 DT 06-NOV-1998 (first entry)

DE Methyl tetrahydropteroyltriglutamate-homocysteine methyltransferase.
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis.
 OS Streptococcus pneumoniae.
 PN WO9743303-A1.
 PD 20-NOV-1997.
 PR 14-MAY-1997; U07950.
 PR 14-MAY-1996; US-017670.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Stodola RK;
 DR WPI: 98-008793/01.
 DR N-PSDB: T98621.
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 PS Claim 12; Pages 342-344; 483pp; English.
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with an E. coli protein, is a methyl tetrahydropteroyltri
 CC glutamate-homocysteine methyltransferase, and is encoded by a DNA of the
 CC invention. The DNA sequences were isolated from S. pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 CC Sequence 456 AA;
 SQ

Query Match 3.4%; Score 116; DB 34; Length 456;
 Best Local Similarity 22.3%; Pred. No. 3.08e+00;
 Matches 35; Conservative 43; Mismatches 68; Indels 11; Gaps 10;

Db 246 vrd-vyadiivlpvdaigl-dfvegkktlelykgafpdkltlyqvinknwnnyeks 303
 QY 154 VPDHYKIIDALPVIAHPMTQFATGVMAIQ-VQSEF--QKAYERGIHKS-KLWEPYEDS 209
 Db 304 lavleqip-aenivltsscllh-vpftaneefepalnhfafavekldeirdl-dair 360
 QY 210 MSLIAQVPLVAAYVYRRMYKNGNTIP-KDDSLDYGANFAHMLGFSSDMHELMKLYTIIH 268
 Db 361 ngqgsealaank-elfatervgenaelrariagldta 396
 QY 269 SDHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAGP 305

RESULT 7
 ID R47340 standard; Protein: 220 AA.
 AC R47340:
 DT 01-JUL-1994 (first entry)
 DE Fragment of chromaffin granule amine transporter protein.
 KW Vesicle membrane transport protein; gene therapy; screening;
 KW Parkinsons disease; neurotoxin; identification; detection;
 KW antibody; probe; chromaffin granule amine transporter protein.
 OS Rattus rattus.
 PN WO9325699-A.
 PD 23-DEC-1993.
 PR 11-JUN-1993; U05704.
 PR 11-JUN-1992; US-899074.
 PR 30-JUL-1992; US-923096.

PA (REGC) UNIV CALIFORNIA.
 PI Edwards RH;
 DR WPI: 94-007556/01.
 PT New mammalian vesicle membrane transport protein - and corresp.
 PT DNA, vectors, transformed cells and antibodies, for diagnosis and
 PT treatment of neurological disorders, e.g. Parkinson disease
 PS Example 2; Page 114-115; 181pp; English.
 CC The cDNA encoding the chromaffin granule amine transporter protein
 CC is useful in gene therapy and as a probe for detecting genomic
 CC sequences. The protein is used for screening cytotoxic compounds
 CC implicated in Parkinsons disease, diseases associated with activity
 CC of neurotoxins or psychiatric disorders and to identify compounds
 CC which selectively inhibit or activate its action. Antibodies raised
 CC against this protein are useful as immunoassay reagents for
 CC detecting the protein and as affinity reagents for purification.
 CC This fragment of the transporter protein shares homology with the
 CC N-terminal domains of the methlenomycin; tetracycline and multi-drug
 CC resistance transporter proteins.
 SQ Sequence 220 AA;

Query Match 3.1%; Score 106; DB 9; Length 220;
 Best Local Similarity 21.1%; Pred. No. 1.47e-01;
 Matches 23; Conservative 40; Mismatches 40; Indels 6; Gaps 6;

Db 24 dnmlltvvvpivptflyatefkdnsnslhrgpsvssqeenvrigilfaskalmqllvnpf 83
 QY 208 DMSLSIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANFAHML-LCFSSD-MHELMKLY 264
 Db 84 vgpbltnrighypmfvgfmlmf-lstlmfafsgtyallfvartllqigs 131
 QY 265 VTTHSDHEGNGVSAHTGHLVASALSDPYLSFAAALNGL-AG-PLHGLAN 311

RESULT 8
 ID W55330 standard; Protein: 329 AA.
 AC W55330:
 DT 15-JUN-1998 (first entry)
 DE H. pylori ORF hp3el1024orf49 protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN WO9737044-A1.
 PD 09-OCT-1997.
 PR 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB: V24739.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claim 14; Pages 556-557; 1145pp; English.
 CC This sequence is a H. pylori protein of unspecified function.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from *H. pylori* by PCR
 CC amplification for recombinant polypeptide production, e.g. in *E. coli*
 CC hosts.
 SQ Sequence 329 AA;

Query Match 3.0%; Score 103; DB 29; Length 329;
 Best Local Similarity 30.6%; Pred. No. 2,33e+01;
 Matches 11; Conservative 15; Mismatches 8; Indels 2; Gaps 2;
 Db 218 veempliasvifnrl-kkmpqlmqdgalny-qefsh 251
 QY 213 IAOVPLVAAYVYRMKYKNGTIPKDDSLDYGANFAH 248

RESULT 9

ID W55574 standard; Protein: 329 AA.
 AC W55574;
 DT 24-JUN-1998 (first entry)
 DE *H. pylori* ORF 06p30223_23557202_c2_130 cytoplasmic protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN W09737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB; V24983.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent *H. pylori*
 PT infection and for diagnosis of *H. pylori* infection
 PS Claims 14,96; Page 779-779; 1145pp; English.
 CC This sequence is a *H. pylori* cytoplasmic protein involved in cofactor
 CC metabolism. The protein may be used in a vaccine to prevent or treat *H.*
 CC *pylori* infection or to identify *H. pylori* polypeptide binding compounds,
 CC useful as potential *H. pylori* life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC *H. pylori* in a sample and the diagnosis of *H. pylori* infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of *H. pylori* mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of *H. pylori*-specific antigens. The genomic sequence of
 CC *H. pylori* (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely *H. pylori* antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from *H. pylori* by PCR
 CC amplification for recombinant polypeptide production, e.g. in *E. coli*
 CC hosts.
 SQ Sequence 329 AA;

Query Match 3.0%; Score 103; DB 29; Length 329;
 Best Local Similarity 30.6%; Pred. No. 2,33e+01;
 Matches 11; Conservative 15; Mismatches 8; Indels 2; Gaps 2;

Db 218 veempliasvifnrl-kkmpqlmqdgalny-qefsh 251
 QY 213 IAOVPLVAAYVYRMKYKNGTIPKDDSLDYGANFAH 248

RESULT 10

ID W29179 standard; Protein: 355 AA.
 AC W29179;
 DT 19-DEC-1997 (first entry)
 DE Rat CC chemokine receptor.
 KW Rat; CC chemokine receptor; screen; binding; ligand.
 OS Rattus rattus.
 PN J09227599-A.
 PD 02-SEP-1997.
 PF 22-FEB-1996; 035192.
 PR 22-FEB-1996; JP-035192.
 PA (TAKE) TAKEDA CHEM IND LTD.
 DR WPI: 97-486426/45.
 DR N-PSDB; T86839.
 PT CC chemokine receptor protein - useful to screen for novel binding
 PT compounds
 PS Claim 1; Page 20-21; 26pp; Japanese.
 CC This sequence is a rat CC chemokine receptor. The receptor can be used
 CC to screen for novel binding compounds and for preparation of antibodies
 CC or antiserum.
 SQ Sequence 355 AA;

Query Match 3.0%; Score 102; DB 24; Length 355;
 Best Local Similarity 29.5%; Pred. No. 2,71e+01;
 Matches 26; Conservative 17; Mismatches 38; Indels 7; Gaps 7;
 Db 73 lfnlavsdvlfvltfpwidy-kldnvwfvgdamckllsgfyvlglyseifillitdr 131
 QY 379 LFOL-VAKLYEVF-LQFLONLAKL-NPWPVNDVHSGVLLN-VY-GLTEARYVTVLFCVSR 433
 Db 132 ylaivhavfslrartvfgiltsiii-w 158
 QY 434 ALGICSLIWDRLGLPLERPKSVTNEW 461

RESULT 11

ID W50891 standard; Protein: 3084 AA.
 AC W50891;
 DT 07-DEC-1998 (first entry)
 DE Mouse laminin A chain.
 KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gertmann-Strausler syndrome; kuru; scrapie; haemodialysis;
 KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
 KW therapy.
 OS Mus Sp.
 FH Key
 FT Domain Location/Qualifiers
 FT 2746..2922
 FT Region /note= "fourth globular domain repeat (Claim 13)"
 FT 2690..2700
 FT /note= "beta-amyloid protein binding region
 (Claim 12)"

PN W09815179-A1.
 PD 16-APR-1998.
 PF 08-OCT-1997; U18145.
 PR 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 DR WPI: 98-240534/21.
 PT Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CJD
 PS Claim 15; Page 74-79; 132pp; English.
 CC This is the amino acid sequence of the mouse laminin A chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidoses. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (mercosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and

CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidosis such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-association amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gertsmann-Straussler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polyneuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 SQ Sequence 3084 AA;

Query Match 2.8% Score 101; DB 35; Length 3084;
 Best Local Similarity 26.7% Pred. No. 3.15e+01;
 Matches 24; Conservative 26; Mismatches 33; Indels 7; Gaps 7;
 Db 1884 selqragaldrdlen-vrvvsnatsaahvshnigtlt-eeawlaa-da-hkntanktd 1939
 QY 11 KLSRSRA-VQQTNLNSVRVQLVQTSGGLRSELOELIPEQQDRKLKSEHGKVLQGN 69
 Db 1940 slse-slas-rqkavlrqrfrfkesvgtr 1967
 QY 70 ITVDVVLGGMRGTMGLWETSLLDPDEGR 99

RESULT 12
 ID R52749 standard; Protein; 355 AA.
 AC R52749; 1995 (first entry)
 DE C-C chemokine receptor.
 KW C-C CKR-1; cytokine; inflammation.
 OS Homo sapiens.
 PN W09411504-A.
 PD 26-MAY-1994.
 PF 04-NOV-1993; U10672.
 PR 10-NOV-1992; US-974025.
 PA (GETH) GENENTECH INC.
 PI Horuk R, Neote K, Schall T;
 DR WPI: 94-183505/22.
 DR N-PSDB: Q62695.
 PT New C-C chemokine receptor and nucleic acid - are used to develop
 PT prods. for use in diagnosis and therapy of inflammation and other
 PT cytokine-mediated disorders
 PS Claim 1; Fig 9; 90pp; English.
 CC The sequence is that of the C-C chemokine receptor. The sequence can
 CC be used in therapeutic or diagnostic compsns. for inflammation and
 CC other cytokine mediated disorders.
 CC See also R52750-2.
 SQ Sequence 355 AA;

Query Match 2.8% Score 98; DB 11; Length 355;
 Best Local Similarity 30.6% Pred. No. 4.95e+01;
 Matches 22; Conservative 15; Mismatches 29; Indels 6; Gaps 6;
 Db 65 lkmtstyllnlaisdlflftlpfwdy-klkddwvfgdamckilsgfytygyseliff 123
 QY 371 LKHLPEPLFOL-VAKLYEYF-LQFLQNLAKL-NPWPNVDAHSGVLLN-YY-GLTEARYY 425

Db 124 illitidrylai 135
 QY 426 TVLFQVSRALGI 437
 RESULT 13
 ID W25751 standard; Protein; 355 AA.
 AC W25751; 1997 (first entry)
 DT 20-NOV-1997
 DE Human MIP-1-alpha/RANTES receptor protein.
 KW Human MIP-1 alpha/RANTES receptor; osteoporosis; pCCR; digestive ulcer;
 KW macrophage inflammatory protein 1 alpha; diabetes; central disease;
 KW regulated on activation, normal T cell expressed and secreted; allergy;
 KW affinity compound; expression vector; CHO cell; viral disease;
 KW infectious disease; tumour; hyperlipidaemia; hypercholesterolaemia.
 OS Homo sapiens.
 PN J09176048-A.
 PD 08-JUL-1997.
 PF 28-DEC-1995; 342130.
 PR 28-DEC-1995; JP-342130.
 PA (TAKE) TAKEDA CHEM IND LTD.
 DR WPI: 97-399449/37.
 DR N-PSDB: T86154.
 PT Preparation of human MIP-1-alpha/RANTES receptor protein - used in
 PT the treatment of viral diseases, tumours, allergy, diabetes
 PT osteoporosis etc.
 PS Disclosure; Page 15-16; 19pp; Japanese.
 CC This sequence represents human MIP-1 alpha/RANTES receptor (macrophage
 CC inflammatory protein 1 alpha/regulated on activation, normal T cell
 CC expressed and secreted). The human MIP-1 alpha/RANTES receptor protein
 CC may be used in a method for the screening of human MIP-1 alpha/RANTES
 CC receptor affinity compounds. The MIP-1 alpha/RANTES receptor coding
 CC sequence may be included in an expression vector, preferably pCCR, and
 CC used to transform a CHO cell for use in the same method. The receptor
 CC protein can provide a preventive and treating agent for viral diseases,
 CC infectious diseases, tumours, allergy, diabetes, central diseases,
 CC hyperlipidaemia, hypercholesterolaemia, osteoporosis, digestive ulcers,
 CC etc.
 SQ Sequence 355 AA;

Query Match 2.8% Score 98; DB 24; Length 355;
 Best Local Similarity 30.6% Pred. No. 4.95e+01;
 Matches 22; Conservative 15; Mismatches 29; Indels 6; Gaps 6;
 Db 65 lkmtstyllnlaisdlflftlpfwdy-klkddwvfgdamckilsgfytygyseliff 123
 QY 371 LKHLPEPLFOL-VAKLYEYF-LQFLQNLAKL-NPWPNVDAHSGVLLN-YY-GLTEARYY 425
 Db 124 illitidrylai 135
 QY 426 TVLFQVSRALGI 437
 RESULT 14
 ID W26588 standard; Protein; 355 AA.
 AC W26588;
 DT 21-JAN-1998 (first entry)
 DE Human MIP-1 alpha/RANTES receptor.
 KW Macrophage inflammatory protein-1 alpha; MIP-1 alpha;
 KW reduced upon activation normal T expressed and secreted; RANTES;
 KW receptor; cytokine; antiinflammatory; inflammation; human.
 OS Homo sapiens.
 PN US5652133-A.
 PD 29-JUL-1997.
 PF 28-JAN-1993; 012988.
 PR 28-JAN-1993; US-012988
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Murphy PW;
 DR WPI: 97-392945/36.
 DR N-PSDB: T90384.
 PT MIP-1-alpha and RANTES receptor nucleic acid - used to develop
 PT products for the detection of these cytokine(s) and their receptors,
 PT particularly in inflammatory processes

PS Claim 2; Column 15-18; 12pp; English.
 CC This polypeptide comprises a claimed receptor for human macrophage
 CC inflammatory protein-1 alpha (MIP-1 alpha) and regulated upon
 CC activation normal T expressed and secreted (RANTES) protein. Also
 CC claimed are: a nucleic acid (see 190384) that encodes the receptor;
 CC a subsequence of the nucleic acid, having at least 12 contiguous
 CC nucleotides; a cell transformed or transfected with the nucleic
 CC acid; and purified MIP-1 alpha/RANTES receptor polypeptide. The
 CC products can be used for detecting the MIP-1 alpha/RANTES receptor
 CC and polymorphisms in physiological samples. In addition, the
 CC receptor can be expressed and used to assay for MIP-1a/RANTES in
 CC biological samples. The quantitation of MIP-1 alpha/RANTES is
 CC useful for monitoring the levels of these cytokines in a patient.
 CC Such measurements are useful in following the antiinflammatory
 CC effects of drugs and prospective usefulness of new antiinflammatory
 CC agents.
 SQ Sequence 355 AA;

Query Match 2.8%; Score 98; DB 25; Length 355;
 Best Local Similarity 30.6%; Pred. No. 4.95e+01;
 Matches 22; Conservative 15; Mismatches 29; Indels 6; Gaps 6;

Db 65 lknmtsiylnlnaisdlflfplfwidy-klkddwvfgdamckilsgfyvtglyseiff 123
 QY 371 LKHLPEDPQL-VAKLYEVF-LQFLQNLAKL-NPWPNVDAHSGVLLN-YY-GUTEARYY 425
 Db 124 illttdrylai 135
 QY 426 TVLFGVSRLGI 437

RESULT 15
 ID P81045 standard; Protein; 2164 AA.
 AC P81045;
 DT 21-AUG-1991 (first entry)
 DE Sequence of the viral proteins VP1-VP4, P2A-P2C, PCA-P3C encoded by
 DE the genomic RNA of rhinovirus strain HRV89
 KW Vaccine: diagnosis.
 OS Rhinovirus.

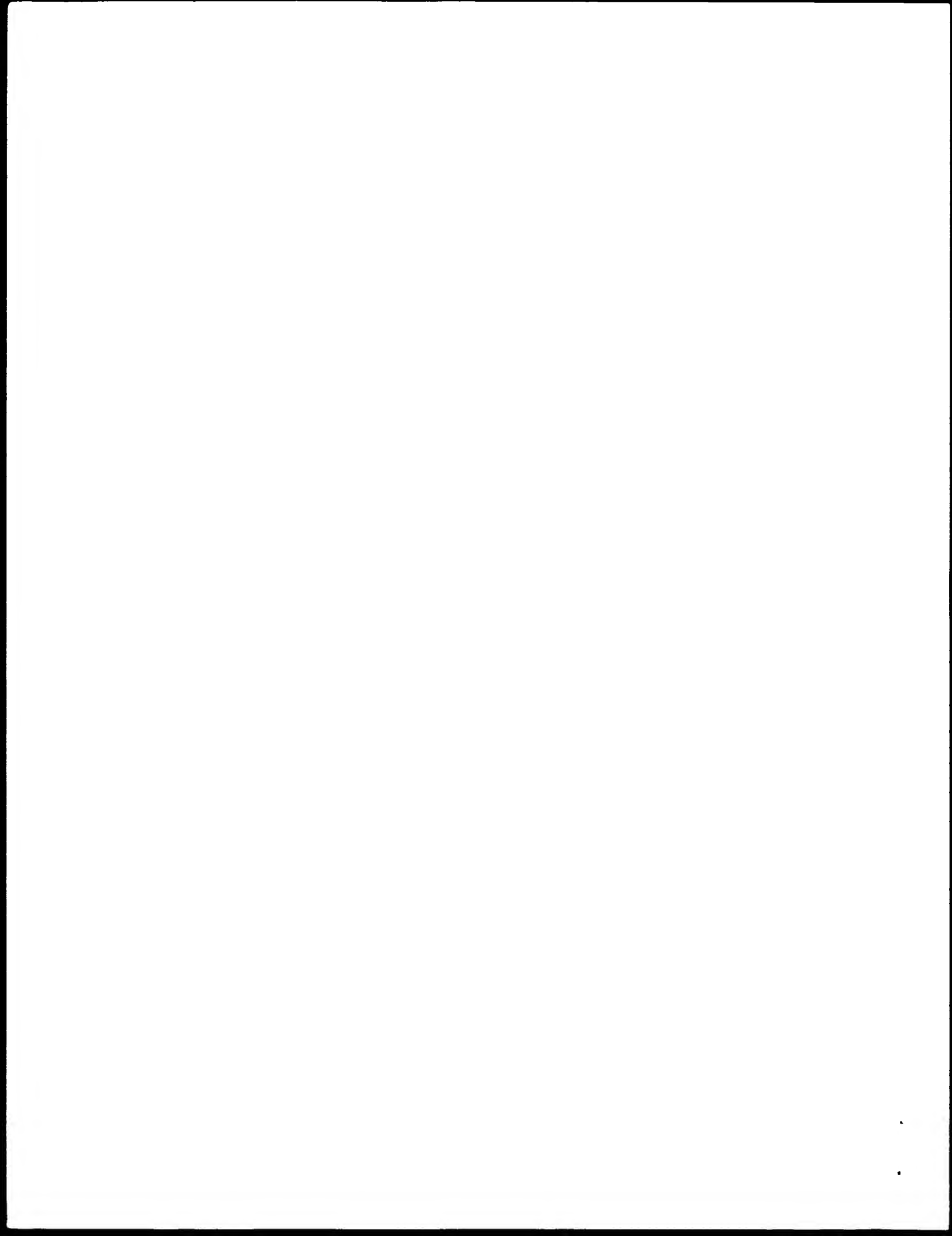
FH Key Location/Qualifiers
 FT protein 1..69

FT /label= VP4
 FT protein 70..337
 FT /label= VP2
 FT protein 338..575
 FT /label= VP3
 FT protein 576..872
 FT /label= VP1
 FT protein 873..1007
 FT /label= P2-A
 FT protein 1008..1101
 FT /label= P2-B
 FT protein 1102..1421
 FT /label= PC-2
 FT protein 1422..1496
 FT /label= P3-A
 FT protein 1497..1517
 FT /label= VPg
 FT protein 1518..1678
 FT /label= PROTEASE
 FT protein 1679..2163
 FT /label= POLYMERASE

PN EP-261403-A.
 PD 30-MAR-1988.
 PF 20-AUG-1987; 112104.
 PR 23-AUG-1986; DE-628658.
 PR 17-JAN-1987; DE-701301.
 PA (BOEH) Boehringer Ingelheim.
 PI Duechler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas,
 PI Kuchler E, Fraseli L, Zorn M;
 DR WP1: 88-085735/13.
 DR N-PSDB: N81390.
 PT New DNA corresponding to viral RNA of rhino-virus HRV89 - useful for

PT prodn. of polypeptide(s) for stimulating immune system against HRV
 PT 89.
 PS Claim 8; Fig 4: 66pp; German.
 CC The viral proteins are used for stimulating a protective immune
 CC response and for blocking cellular receptors. Abs against them
 CC are useful for assay and purificn. of the corresp. antigen, and
 CC can also be used for the therapeutic and diagnostic applications.
 SQ Sequence 2164 AA;
 Query Match 2.8%; Score 95; DB 2; Length 2164;
 Best Local Similarity 40.5%; Pred. No. 7.73e+01;
 Matches 15; Conservative 11; Mismatches 9; Indels 2; Gaps 2;
 Db 1538 vvttdkg-ktgigdydqvmlphtshdpgseilvdgv 1573
 QY 91 LLDPEGIRFRLSIYECQKVLPA-AKPGGEPLPEGL 126

Search completed: Fri Oct 22 17:04:01 1999
 Job time : 253 secs.



CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Robbins, Berliner & Carson
CC STREET: 201 N. Figueroa Street, 5th Floor
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: USA
CC ZIP: 90012-2628
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/00476
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Berliner, Robert
CC REGISTRATION NUMBER: 20,121
CC REFERENCE/DOCKET NUMBER: 5555-291
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 310-977-1001
CC TELEFAX: 310-977-1003
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 355 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
CC
CC SEQUENCE 355 AA; 41172 MW; 710742 CN;

Query Match 2.8%; Score 98; DB 3; Length 355;

Best Local Similarity 30.6%; Pred. No. 2.58e+01;

Matches 22; Conservative 15; Mismatches 29; Indels 6; Gaps 6;

Db 65 LKNMTSYLLNLAISDLFLTPWIDY-KLKDDWVFGDAMCKILSGFYTTGLYSIFF 123

QY 371 LKHPEDPLFQL-VAKLYEVF-LQFLQNLAKL-NPWPVNDHSGVLLN-YY-GLTEARYY 425

Db 124 IILLTIDRYLAI 135

QY 426 TVLFGVSRALGI 437

RESULT 7
ID US-08-756-317-9 STANDARD; PRT: 624 AA.

XX
AC
XX
DT
DT
XX

Sequence 9, Application US/08756317

DE Sequence 9, Application US/08756317

CC Patent No. 5849894

CC GENERAL INFORMATION:

CC APPLICANT: Clemente, Thomas E.

CC APPLICANT: Kishore, Ganesh M.

CC APPLICANT: Mitsky, Timothy A.

CC APPLICANT: Stark, David M.

CC TITLE OF INVENTION: Improved Rhodospirillum Rubrum

CC TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase

CC NUMBER OF SEQUENCES: 15

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Arnold, White & Durkee

CC STREET: P.O. Box 4433

CC CITY: Houston

CC STATE: TX

CC COUNTRY: USA

CC ZIP: 77210-4433

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/756,317
CC FILING DATE: 25-NOV-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/007,693
CC FILING DATE: 29-NOV-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Patterson, Melinda L.
CC REGISTRATION NUMBER: 33,062
CC REFERENCE/DOCKET NUMBER: MOBT:008
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (713) 787-1400
CC TELEFAX: (713) 787-1440
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 624 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC SEQUENCE 624 AA; 69538 MW; 1929028 CN;

Query Match 2.8%; Score 97; DB 2; Length 624;

Best Local Similarity 27.1%; Pred. No. 2.99e+01;

Matches 23; Conservative 22; Mismatches 33; Indels 7; Gaps 7;

Db 90 WIKLLAQO-PERVIGQOV-SY-WGETLHFAEAQAFAFARGITVPPPSGPRDRRFA-NPL 145

QY 317 WIKSVVECGENISKEQLKYAW-KTLKS-GKVVPGFGHGVLRKTDPRYTCQREFALKHL 374

Db 146 WEAHFFNFIRKQYQINAAQLEAA 170

QY 375 -PEDPLFOLVAKLYEVFLQFLQNL 398

RESULT 8

ID US-08-465-273-14 STANDARD; PRT: 80 AA.

XX
AC
XX
XX
DT
DT
XX

Sequence 14, Application US/08465273

CC Sequence 14, Application US/08465273

CC Patent No. 5866361

CC GENERAL INFORMATION:

CC APPLICANT: Choulaka, Andre

CC APPLICANT: Perrin, Arnaud

CC APPLICANT: Dujon, Bernard

CC APPLICANT: Nicolas, Jean-Francois

CC TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme

CC Patent No. 5866361

CC TITLE OF INVENTION: I-SCEI and the Uses Thereof

CC NUMBER OF SEQUENCES: 52

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

CC ADDRESSEE: Dunner

CC STREET: 1300 I Street, N.W.

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20005-3315

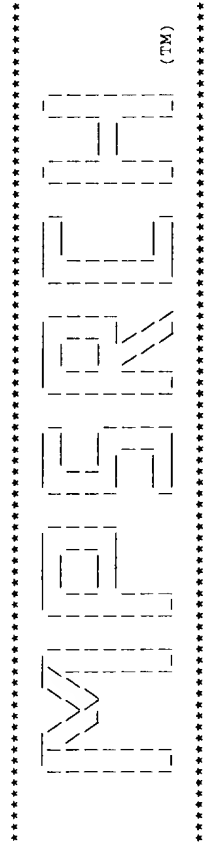
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25



Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 17:04:20 1999; MasPar time 20.94 Seconds
Tabular output not generated. 897.479 Million cell updates/sec

Title: >US-08-702-718-6
Description: (1-469) from US08702718.pep
Perfect Score: 3443
Sequence: 1 MVFYRGVSLLSKLRRAVQO.....PLERPKSVTMEWLENHCKKA 469

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 50.451; Variance 107.716; scale 0.468

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3098	90.0	471	2	S44316 ethanolamine ammonia-	0.00e+00
2	2987	86.8	474	2	T02390 probable citrate (si)	0.00e+00
3	2454	71.3	472	1	YKUM citrate (si)-synthase	0.00e+00
4	2021	58.7	464	1	YKPG citrate (si)-synthase	0.00e+00
5	1820	52.9	468	2	S42370 citrate (si)-synthase	0.00e+00
6	1763	51.2	479	1	YKBY citrate (si)-synthase	0.00e+00
7	1679	48.8	460	1	YKBYC citrate (si)-synthase	1.58e-293
8	1680	48.8	469	2	S41563 citrate (si)-synthase	1.00e-293
9	1385	40.2	462	2	JC5625 14-nm filament protei	1.29e-235
10	1218	35.4	486	2	S52814 citrate (si)-synthase	6.61e-203
11	360	10.5	437	2	I40717 citrate (si)-synthase	2.58e-40
12	347	10.1	373	2	A43936 citrate synthase - Ba	4.98e-38
13	310	9.0	436	1	YKQC citrate (si)-synthase	1.33e-31
14	305	8.9	372	2	E64658 citrate synthase III	9.59e-31
15	306	8.9	397	2	S74344 citrate synthase gl-A	6.45e-31
16	287	8.3	428	2	I39506 citrate (si)-synthase	1.12e-27
17	285	8.3	428	1	YKFSCA citrate (si)-synthase	2.45e-27
18	274	8.0	430	2	JQ1392 citrate (si)-synthase	1.75e-25
19	275	8.0	431	2	E70782 probable gltA2 protei	1.19e-25
20	272	7.9	372	2	C69417 citrate synthase (cit	3.79e-25
21	271	7.9	516	2	S53007 citrate synthase - cu	5.58e-25
22	261	7.6	426	2	D71982 citrate synthase - He	2.63e-23
23	248	7.2	372	2	G69600 citrate synthase II	3.55e-21

ALIGNMENTS

RESULT 1

ENTRY S44316 #type complete
TITLE ethanolamine ammonia-lyase (EC 4.3.1.7) - potato
ORGANISM #formal_name Solanum tuberosum common_name potato
DATE 08-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 08-Sep-1997

ACCESSIONS S44316
REFERENCE S44316
#authors Landschuetz, V.; Willmitzer, L.; Mueller-Roeber, B.
#submission Submitted to the EMBL Data Library, September 1993
#accession S44316

##molecule_type mRNA
##residues 1-471 ##label LAN
##cross-references EMBL:X75082; NID:q483509; PID:q483510
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS ammonia-lyase; carbon-nitrogen lyase
SUMMARY #length 471 #molecular-weight 52612 #checksum 4730

Query Match 90.0%; Score 3098; DB 2; Length 471;
Best Local Similarity 89.6%; Pred. No. 0.00e+00;
Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4;

Db	1	MVFYRGVSLLSKLRRAVQO	SNVSRVWLVQVOTSSGLDLRSELVQELIPEQODRLKKIK	60
Qy	1	MVFYRGVSLLSKLRRAVQO	TNLSNVSRYWLVQVOTSSGLDLRSELVQELIPEQODRLKKIK	59
Db	61	SDM-KGSIGNITVDMVLGGMRG	MTGLLWPKPHYLDPOEGIRFRGLSIPECOKVLPAAPGG	119
Qy	60	SEHCKVOLGNITVDMVLGGMRG	MTGLLWETSLDDPEGIRFRGLSIPECOKVLPAAPGG	119
Db	120	EPPEGLLWLLTGKVPKSEQVNS	IVSGISAEISLIIMYTTIDALPVTAHPTQFATG	179
Qy	120	EPLPEGLLWLLTGKVPKSEQVNS	IVSGISAEISLIIMYTTIDALPVTAHPTQFATG	178
Db	180	VMAQVQSEFQKAYEKG	IHKSWEPTYEDSMNLIAQVPLVAAVYVRYKNGDITPKDE	239
Qy	179	VMAQVQSEFQKAYEKG	IHKSWEPTYEDSMNLIAQVPLVAAVYVRYKNGDITPKDD	238
Db	240	SLDYGANFAHMLG	FSSEMHELLMRLVYVTHSDHEGNSAHTCHLVASALSDFLSFAA	299
Qy	239	SLDYGANFAHMLG	FSSEMHELLMRLVYVTHSDHEGNSAHTCHLVASALSDFLSFAA	297
Db	300	ALNGLAGPLHGLANQ	EVLMIKSVBECGNSISKEQLKDYVWTKLNSGKVPFGHGVLR	359
Qy	298	ALNGLAGPLHGLANQ	EVLMIKSVBECGNSISKEQLKDYVWTKLNSGKVPFGHGVLR	357
Db	360	KTVPRYTCQPEFAMK	HLPEDPQLVSKLVEVFLFLFLQNLAKLKPWPNDVDAHSSVLLNY	419

```

QY 358 KTDPRYTCOREFALKHLPEDPLFQLVAKLYEVFLQFQNLAKLPNPNVDAHSGVLLNY 417
|||||
Db 420 GLTEARYTYTLFGVSRAIGCSQLIWDRLALGLPLERPKSVTMWLENCKKA 471
|||||
QY 418 GLTEARYTYTLFGVSRAIGCSQLIWDRLALGLPLERPKSVTMWLENCKKA 469
|||||

RESULT 2
ENTRY #type complete
TITLE probable citrate (si)-synthase (EC 4.1.3.7) - Arabidopsis
thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
05-Mar-1999
ACCESSIONS T02390
REFERENCE 214177
#authors Rounslev, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.P.; Venter,
J.C.
#submission submitted to the EMBL Data Library, May 1998
#description Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
#accession T02390
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-474 ##label ROU
##cross-references EMBL:AC004521; NID:g3128166; PID:g3128180
GENETICS
#map_position 2
#introns 17/3; 39/3; 53/3; 75/3; 97/3; 135/3; 157/1; 184/3; 201/2;
227/2; 270/2; 284/3; 313/3; 347/3; 383/3; 400/3; 425/2;
442/3
#note F411.16
#KEYWORDS carbon-carbon lyase; oxo-acid-lyase; tricarboxylic acid cycle
#SUMMARY #length 474 #molecular-weight 52782 #checksum 6960

Query Match 86.8%; Score 2987; DB 2; Length 474;
Best Local Similarity 82.6%; Pred. No. 0.00e+00;
Matches 389; Conservative 53; Mismatches 27; Indels 2; Gaps 2;

Db 1 MYFFRSVSAFTLRSRVQGOSSLSNSVRWQMOSSTDLDKLSQLOELIPEQODRLKLLK 60
|||||
QY 1 MYFYGVSLLSKLSRA-VQOTNLSNSVRWLMQVOTSSGLDLRSELQELIPEQODRLKLLK 59
|||||
Db 61 SEHGKVLGNITVDWVIGGMRGMTGLLWETSLDPEVFGALGDCRLPECOKALPTAQSG 120
|||||
QY 60 SEHGKVLGNITVDWVIGGMRGMTGLLWETSLDPEVFGALGDCRLPECOKALPTAQSG 119
|||||
Db 121 EPLPEGLLWLLTGKVPKSEQVDSLSQELRSRAIVPDHYTKTIDALPVAHPMTQFA 180
|||||
QY 120 EPLPEGLLWLLTGKVPKSEQVDSLSQELRSRAIVPDHYTKTIDALPVAHPMTQFA 179
|||||
Db 181 MALQVQSEFOKAYENGHKSKEWPEPTYEDCLNLIARVPVVAAYVYRMKNGSDIPS 240
|||||
QY 180 MALQVQSEFOKAYENGHKSKEWPEPTYEDCLNLIARVPVVAAYVYRMKNGSDIPS 239
|||||
Db 241 LDYGANFHLGFDDEKVKELMRLYITIHSDHEGNGVSAHTGHLVGSALSDPYLSFAAL 300
|||||
QY 240 LDYGANFHLGFDDEKVKELMRLYITIHSDHEGNGVSAHTGHLVGSALSDPYLSFAAL 299
|||||
Db 301 NGLAGPLHGLANQEVLLWIKSVVEECGEDIKSQELKVEYVMTKNSGKVIIPGYGHVLRNT 360
|||||
QY 300 NGLAGPLHGLANQEVLLWIKSVVEECGEDIKSQELKVEYVMTKNSGKVIIPGYGHVLRNT 359
|||||
Db 361 DPRYVCOREFALKHLPDDPLFQLVAKLYEVFLQFQNLAKLPNPNVDAHSGVLLNY 420
|||||
QY 360 DPRYVCOREFALKHLPDDPLFQLVAKLYEVFLQFQNLAKLPNPNVDAHSGVLLNY 418
|||||
Db 421 LTEARYTYTLFGVSRAIGCSQLIWDRLALGLPLERPKSVTMWLENCKKA 471
|||||

```

```

QY 419 LTEARYTYTLFGVSRAIGCSQLIWDRLALGLPLERPKSVTMWLENCKKA 469
|||||

RESULT 3
ENTRY #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor, mitochondrial -
Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
29-May-1998
ACCESSIONS JA0149; S07004
REFERENCE JA0149
#authors Unger, E.A.; Hand, J.M.; Cashmore, A.R.; Vasconcelos, A.C.
#journal Plant Mol. Biol. (1989) 13:411-418
#title Isolation of a cDNA encoding mitochondrial citrate synthase
from Arabidopsis thaliana.
#cross-references MUID:91370823
#accession JA0149
#molecule_type mRNA
#residues 1-472 ##label UNG
##cross-references EMBL:X17528; NID:g11243; PID:e1188578; FID:g2652924
Fig. 1 in having 47-Gly, 125-Trp, 141-Leu, 151-Ser,
185-Asn, 187-Trp, 191-Asn, 348-Leu, an additional Gly
after 89-Trp, an additional Val after 393-Cys and two
additional Arg after 228-Arg, in lacking 114-Leu and
residues 233 to 241
GENETICS
#genome nuclear
#COMPLEX homodimer
#CLASSIFICATION superfamily citrate (si)-synthase
#KEYWORDS carbon-carbon lyase; homodimer; mitochondrion;
oxo-acid-lyase; tricarboxylic acid cycle
#FEATURE
343-375 #region acetyl-CoA binding #status predicted
309,355,407 #active_site His, His, Asp #status predicted
#SUMMARY #length 472 #molecular-weight 52941 #checksum 9626

Query Match 71.3%; Score 2454; DB 1; Length 472;
Best Local Similarity 71.7%; Pred. No. 0.00e+00;
Matches 339; Conservative 62; Mismatches 64; Indels 8; Gaps 8;

Db 1 MYFFRSVSAFTLRSRVQGOSSLSNSVRWQMOSSTDLDKLSQLOELIPEHDKRLKLLK 60
|||||
QY 1 MYFYGVSLLSKLSRA-VQOTNLSNSVRWLMQVOTSSGLDLRSELQELIPEQODRLKLLK 59
|||||
Db 61 SEHGKVLGNITVDWVIGGMRGMTGLLWETSLDPEVFGALGDCRLPECOKALPTAQSG 120
|||||
QY 60 SEHGKVLGNITVDWVIGGMRGMTGLLWETSLDPEVFGALGDCRLPECOKALPTAQSG 118
|||||
Db 121 GLNHYRSFVASLWKGTL-AKSKLKH-CRKTWNRAVSDYVYNAIDALPSTAHPTQFA 178
|||||
QY 119 G-EPLPEGLL-WLLTGKVPKSEQVDSLSQELRSRAIVPDHYTKTIDALPVAHPMTQFA 176
|||||
Db 179 SGVMAQVQSEFOKAYENGHKSKEWPEPTYEDCLNLIARVPVVAAYVYRMKNGSDIPS 238
|||||
QY 177 TGVMAQVQSEFOKAYENGHKSKEWPEPTYEDCLNLIARVPVVAAYVYRMKNGSDIPS 236
|||||
Db 239 DKSLDYGANFHLGFDDEKVKELMRLYITIHSDHEGNGVSAHTGHLVGSALSDPYLSFA 298
|||||
QY 237 DKSLDYGANFHLGFDDEKVKELMRLYITIHSDHEGNGVSAHTGHLVGSALSDPYLSFA 296
|||||
Db 299 AALNGLAGPLHGLANQEVLLWIKSVVEECGEDIKSQELKVEYVMTKNSGKVIIPGYGHVLRNT 358
|||||
QY 297 AALNGLAGPLHGLANQEVLLWIKSVVEECGEDIKSQELKVEYVMTKNSGKVIIPGYGHVLRNT 356
|||||
Db 359 RNTDPRYVCOREFALKHLPDDPLFQ-CCKLMKL-ASCLTELESEEPWPVNDHSGVLLNH 416
|||||
QY 357 RNTDPRYVCOREFALKHLPDDPLFQ-LVAKLYEVFLQFQNLAKLPNPNVDAHSGVLLNY 416
|||||
Db 417 YGLTEARYTYTLFGVSRAIGCSQLIWDRLALGLPLERPKSVTMWLENCKKA 469
|||||

```

```

QY 417 YGLTEARYTYTLFGVSRALGICSLTWDPALGLPLERPKSVTMEWLENCKKA 469

RESULT 4
ENTRY YKPG #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 15-Oct-1982 #sequence_revision 30-Sep-1992 #text_change
18-Sep-1998

ACCESSIONS A29966; A01109; A61347
REFERENCE A29966
#authors Evans, C.T.; Owens, D.D.; Sumegi, B.; Kispal, G.; Srere, P.A.
#journal Biochemistry (1988) 27:4680-4686
#title Isolation, nucleotide sequence, and expression of a cDNA
encoding pig citrate synthase.
#cross-references MUID:89000665
#accession A29966
#molecule_type mRNA
##residues 1-464 ##label EVA
##cross-references EMBL:M21197; NID:g164418; PID:g164419
REFERENCE A90457
#authors Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Walsh, K.A.;
Titani, K.
#journal Biochemistry (1982) 21:2028-2036
#title Complete amino acid sequence of porcine heart citrate
synthase.
#cross-references MUID:82231993
#accession A01109
#molecule_type protein
##residues 28-464 ##label BLO
REFERENCE A61347
#authors Bloxham, D.P.; Parmelee, D.C.; Kumar, S.; Wade, R.D.;
Ericsson, L.H.; Neurath, H.; Walsh, K.A.; Titani, K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:5381-5385
#title Primary structure of porcine heart citrate synthase.
#accession A61347
#molecule_type protein
##residues 28-464 ##label BL2
REFERENCE A92884
#authors Remington, S.; Wiegand, G.; Huber, R.
#journal J. Mol. Biol. (1982) 158:111-152
#title Crystallographic refinement and atomic models of two
different forms of citrate synthase at 2.7 and 1.7
angstroms resolution.
#cross-references MUID:83010291
#contents annotation: X-ray crystallography, 2.7 and 1.7 angstroms
COMMENT Citrate (si)-synthase is found in nearly all cells capable of
oxidative metabolism. It catalyzes the condensation of
oxaloacetate and acetyl-CoA to form citrate in the tricarboxylic
acid cycle.
COMMENT It is synthesized in the cytoplasm but functions in the
mitochondrion of eukaryote cells.
COMMENT This molecule is a dimer of identical chains. Each dimer binds two
molecules of acetyl-CoA and two molecules of oxaloacetate at two
active sites.
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; homodimer; methylated amino acid;
mitochondrion; oxo-acid-lyase; tricarboxylic acid cycle
FEATURE 1-27
#domain transit peptide (mitochondrion) #s-status
#predicted #label SIG\
#product citrate (si)-synthase #status experimental
28-464
#label MAT\
#active site His, His, Asp #status predicted\
#modified site N6,N6,N6-trimethyllysine (Lys) #status
experimental
SUMMARY #length 464 #molecule-weight 51629 #checksum 3548
Query Match 58.7%; Score 2021; DB 1; Length 464;
Best Local Similarity 61.8%; Pred. No. 0.00e+00;
Matches 265; Conservative 78; Mismatches 84; Indels 2; Gaps 2;
Db 27 SASSTNLKDLADLIPKEQARIKTFQHQHNTVVGQITVDMYGMGMKGLVETSLVD 86

```

```

QY 34 TSSGLDLSRSELQELIPQODRLAKLSEHGKVGOLGNITVDMVLGMRGNTGLLWETSLD 93
Db 87 PDGIRFRGYSIPECOKMLPKAKGGEPLPEGLFWLLVTGQIPTEQVSWLSKEWAKRAA 146
QY 94 PDGIRFPGLSIYECCQKVLPAAPKPGGEPLPEGLFWLLVTGQVPSKEQVDSLSQELSPAT 153
Db 147 LPSHVITMLDNFTNLHPMSQLSAAITALNESFNARAYAEIGHRTKYWELIYEDCMOLI 206
QY 154 VPDHYKTTIDALPTAHPMTQFATGVVALQVQSEFOKAYEKGHIHKLMEPTYEDSMALI 213
Db 207 AKLPCTVAAKIYRMLYREGSGSIGAIDSKLDMSHNFTNMLGYTDAQFTLMRLYITIHSDE 266
QY 214 AQVPLVAAVYRRMYKNGNTI-PKDDSLDYGANFAHMLGFSSSDMHMLKLYVTIHSDE 272
Db 267 GGNVSAHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANOEVLLWLTOLQEKVCKDVSDE 326
QY 273 GGNVSAHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANOEVLLWLTOLQEKVCKDVSDE 332
Db 327 KLRDYIWNLTNSGRVPGYGHAVLRKTDPRYTCOREFALKHLPDPMFKLVLAOLYKIVPN 386
QY 333 QLRDYAWKTLKSGVVPFGHGVLRKTDPRYTCOREFALKHLPDPMFKLVLAOLYKIVPN 392
Db 387 VLLEQGRKAKPNVNDVAHSGVLLQYGYGTMETMYTYTVLVGVSRAQLVLAQLINSGALGFPL 446
QY 393 FLQNLAKL-NPWPNDVAHSGVLLQYGYGTMETMYTYTVLVGVSRAQLVLAQLINSGALGFPL 451
Db 447 ERPKSNSTD 455
QY 452 ERPKSVTME 460

RESULT 5
ENTRY S42370 #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor - Caenorhabditis
elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change
18-Sep-1998
ACCESSIONS S42370
REFERENCE S42368
#authors Smith, A.
#submission submitted to the EMBL Data Library, March 1994
#accession S42370
#status preliminary
#molecule_type DNA
##residues 1-468 ##label SWI
##cross-references EMBL:Z30423; NID:g458479; PID:g458482
GENETICS
#introns 69/3; 202/3; 309/3
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; mitochondrion; oxo-acid-lyase
SUMMARY #length 468 #molecule-weight 51540 #checksum 2659
Query Match 52.9%; Score 1820; DB 2; Length 468;
Best Local Similarity 56.5%; Pred. No. 0.00e+00;
Matches 245; Conservative 82; Mismatches 103; Indels 4; Gaps 4;
Db 25 LSTSAEGSTNLKELKIPAHNAKVKSFRTEHSTGVQNVNIDMYGMRSMGMVET 84
QY 30 LOVOTSSGLDLSRSELQELIPQODRLAKLSEHGKVGOLGNITVDMVLGMRGNTGLLWET 89
Db 85 SVLDPEGIRFRGYSIPECOKMLPKAKGGEPLPEAIWLLCTGDPVSEAOATAITKEWN 144
QY 90 SLDPDDEGIRFRGLSIECCQKVLPAAPKPGGEPLPEGLFWLLVTGQVPSKEQVDSLSQELR 149
Db 145 ARADLPHTVYVMDLNFDPNLHPAQFTAAALNNESKFAGAYARGVAKASYEYAYEDS 204
QY 150 SRATVPDHYKTTIDALPTAHPMTQFATGVVALQVQSEFOKAYEKGHIHKLMEPTYEDS 209
Db 205 MDLLAKLPTVAAYIRNLYRDGSAVSVIDPKDKDWSANFSSMLGYDDPLFAELMLRLVLIH 264
QY 210 MSLLAQVPLVAAVYRRMYKNGNTIP-KDSDLDYGANFAHMLGFSSSDMHMLKLYVTIH 268

```

```

Db 265 SDHEGNSAHTSHLYGSALSOPYLSFSAAMAGLAGPLHGLANOEVLVFLNKLIVGEIGFN 324
QY 269 SDHEGNSAHTSHLYGSALSOPYLSFSAALNGLAGPLHGLANOEVLVFLNKLIVGEIGFN 328
Db 325 YTEQOLKEWKKHKGQVVPVGHAVLRKTDPRVCEQREFALKHLPNDLFLKLYSTLYK 384
QY 329 ISKEQLDYAWKTLKSGVVPVGHGVLKTDPRVCEQREFALKHLPNDLFLKLYSTLYK 388
Db 385 ITPGILLEGCKAKNPVNDHSGVLLQYFGMTMSFYVFLFGVSRALG-CLISQIWAR 443
QY 389 VFLQFLQNLAKL-NPWPVNDHSGVLLQYFGMTMSFYVFLFGVSRALG-CLISQIWAR 446
Db 444 MGLPLERPKSHSTD 457
QY 447 LGLPLERPKSVTME 460

RESULT 6
ENTRY YKBY #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) precursor, mitochondrial - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein N2019; protein YNR001c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 25-Feb-1985 #sequence_revision 10-Feb-1995 #text_change 26-Feb-1999
ACCESSIONS S35390; S45124; S48338; A01110; S63327
REFERENCE S35390
#authors Lindner, P.
#submission submitted to the EMBL Data Library, July 1993
#accession S35390
#molecule_type DNA
#residues 1-479 #label LIN
#cross-references EMBL:23259; NID:g313749; PID:g313750
REFERENCE S45118
#authors Verhasselt, P.; Aert, R.; Voet, M.; Voelckaert, G.
#submission submitted to the EMBL Data Library, January 1994
#description Twelve open reading frames revealed on the 23.6 kbp segment flanking the centromere on the Saccharomyces cerevisiae chromosome XIV right arm.
#accession S45124
#molecule_type DNA
#residues 1-479 #label VER
#cross-references EMBL:X77395; NID:g496717; PID:g496718
REFERENCE S48338
#authors Verhasselt, P.; Aert, R.; Voet, M.; Voelckaert, G.
#journal Yeast (1994) 10:1355-1361
#title Twelve open reading frames revealed in the 23.6 kb segment flanking the centromere on the Saccharomyces cerevisiae chromosome XIV right arm.
#cross-references MUID:95208356
#accession S48338
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-479 #label VE2
#cross-references EMBL:X77395; NID:g496717; PID:g496718
#note the nucleotide sequence was submitted to the EMBL Data Library, January 1994
REFERENCE A01110
#authors Suisse, M.; Suda, K.; Schatz, G.
#journal EMBO J. (1984) 3:1773-1781
#title Isolation of the nuclear yeast genes for citrate synthase and fifteen other mitochondrial proteins by a new screening method.
#cross-references MUID:85003587
#accession A01110
#molecule_type DNA
#residues 1-57, 'Q', 59-77, 'E', 78-479 #label SUI
#cross-references GB:X00782; NID:g3602; PID:g3603
#experimental_source strain D273-10B
REFERENCE S62910
#authors Aert, R.; Verhasselt, P.; Voet, M.; Voelckaert, G.
#submission submitted to the Protein Sequence Database, April 1996

```

```

#accession S63327
#molecule_type DNA
#residues 1-479 #label AER
#cross-references EMBL:271616; NID:g1302468; PID:e233778; PFD:g1302469; MIPS:YNR001c
#experimental_source strain S288C
GENETICS
#gene SGD:CIT1
#cross-references SGD:S0005284; MIPS:YNR001c
#map_position 14R
#genome nuclear
COMPLEX homodimer
FUNCTION
#description carbon-carbon lyase; catalyzes condensation of oxaloacetate and acetyl-CoA to form citrate; oxo-acid-lyase
#pathway tricarboxylic acid cycle
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS acetyl-CoA; carbon-carbon lyase; homodimer; mitochondrion; oxo-acid-lyase; tricarboxylic acid cycle
FEATURE
1-39 #domain transit peptide (mitochondrion) #status predicted #label INP\
40-479 #product citrate (si)-synthase #status predicted #label MAT\
312,358,413 #active_site His, His, Asp #status predicted
SUMMARY #length 479 #molecular_weight 53360 #checksum 4782
Query Match 51.2%; Score 1763; DB 1; Length 479;
Best Local Similarity 55.4%; Pred. No. 0.00e+00;
Matches 237; Conservative 88; Mismatches 102; Indels 1; Gaps 1;
Db 39 SASQTLKERFAEIIIPAKAEIKKFKHCKTVIGVLLFOAYGGMRGKGLVWEGSVLD 98
QY 34 TSSGLDRSELQELIPEQDRLLKLLKSEHGVOLGNITVDMVLGGMRGTLGLWETSLD 93
Db 99 PEEGIRFGRITPEIQRELKAEKSTPELPEALFWLLLTGEIPTDAQKALSADLAARSE 158
QY 94 PDEGIRFGLSIQECQKVLPAKPGGPELPEGLLWLLLTCKVPSKEQVDSLSQELSRAT 153
Db 159 IPEHVIOQLSLPKDLHPMAQFSIANTALESEKFAKAYAGVSKKEYWSYTFEDSLDL 218
QY 154 VPDHVYKTDALPVTAHPMTQFATGVNALQVSEFOKAEKGIHKSKEWPTVEDSMSLI 213
Db 219 GKLPIASKIYRNKFKDKITDTPNADYKGNLAQLLGYENKDFIDMLRLYLTIHSDHEG 278
QY 214 AQVPLVAAYVYRMKNGNIPKDDSLDYCANFAHMLGFSSDMHMLKLYVTHSDHEG 273
Db 279 GNVSATHTLVGSALSPYLSLAAGNLGAGPLHGRANQEVLEWLPKLRREVGYDSKET 338
QY 274 GNVSATHTLVGSALSPYLSLAAGNLGAGPLHGRANQEVLEWLPKLRREVGYDSKET 333
Db 339 IEKYLWDTLWAGVVPVGHVLRKTDPRVTAQREFALKHFPDYLFLKLVSTIYVAPGV 398
QY 334 LKDYAWKTLKSGVVPVGHGVLKTDPRVTAQREFALKHFPDYLFLKLVSTIYVAPGV 393
Db 399 LTRHGKTKNPVNDHSGVLLQYFGMTMSFYVFLFGVSRALG-CLISQIWAR 458
QY 394 LQNLAKL-NPWPVNDHSGVLLQYFGMTMSFYVFLFGVSRALG-CLISQIWAR 452
Db 459 RPKSFSTE 466
QY 453 RPKSVTME 460

RESULT 7
ENTRY YKBYC #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7), peroxisomal - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YCR005c; protein YCR043
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 26-Feb-1999
ACCESSIONS A25393; S26734; S19474; S12944; A39664

```

REFERENCE

#authors Rosenkrantz, M.; Alam, T.; Kim, K.S.; Clark, B.J.; Srere, P.A.; Guarente, L.P.

#journal Mol. Cell. Biol. (1986) 6:4509-4515

#title Mitochondrial and nonmitochondrial citrate synthases in Saccharomyces cerevisiae are encoded by distinct homologous genes.

#cross-references MUID:87089811

#accession A25393

##molecule_type DNA

##residues 1-460 ##label ROS

##cross-references EMBL:Z11113; NID:g3297; PID:g3299; GB:M14685; NID:g171226; PID:g171227

REFERENCE

S22265

#authors Biceau, N.; Fremaux, C.; Hebrard, S.; Menara, A.; Aigle, M.; Crouzet, M.

#journal Yeast (1992) 8:61-70

#title The complete sequence of a 10.8kb fragment to the right of the chromosome III centromere of Saccharomyces cerevisiae.

#cross-references MUID:92254505

#accession S26734

##status translation not shown

##molecule_type DNA

##residues 1-460 ##label BIT

##cross-references EMBL:X59720; NID:g1907116; PID:e264468; PID:g1907148

REFERENCE

S19452

#authors Aigle, M.; Biteau, N.; Crouzet, M.

#submission submitted to the Protein Sequence Database, March 1992

#accession S19474

##molecule_type DNA

##residues 1-460 ##label AIG

##cross-references EMBL:X59720; NID:g1907116; PID:e264468; PID:g1907148; MIPS:YCR005C

REFERENCE

A39664

#authors Liao, X.; Small, W.C.; Srere, P.A.; Butow, P A

#journal Mol. Cell. Biol. (1991) 11:38-46

#title Intramitochondrial functional functions regulate nonmitochondrial citrate synthase (CIT2) expression in Saccharomyces cerevisiae.

#accession S12944

##molecule_type DNA

##residues 1-24 ##label LIA

##cross-references EMBL:M54982; NID:g171228; PID:g171229

GENETICS

#gene SGD:CIT2

##cross-references SGD:S0000598; MIPS:YCR005C

#map_position 3R

CLASSIFICATION

#superfamily citrate (si)-synthase

KEYWORDS

acetyl-CoA; carbon-carbon lyase; homodimer; oxo-acid-lyase; peroxisome; tricarboxylic acid cycle

FEATURE

458-460 #region peroxisome location signal (S-[RKH]-L) motif

293,339,394 #active_site His, His, Asp #status predicted

SUMMARY

#length 460 #molecular-weight 51413 #checksum 12779

Query Match 48.8%; Score 1679; DB 1; Length 460;

Best Local Similarity 52.4%; Pred. No. 1,58e-293;

Matches 229; Conservative 92; Mismatches 115; Indels 1; Gaps 1;

Db 11 NVASYLQSNSSQKTLKERFSSEIYPHIAQDVQFVKERHKTKISDVLLVEQVYGGVRGTPG 70

Qy 25 NSVRWLQVQTSSGLDLRSELQELIPQODRLKLLKSECHKVOLGNITVDWLGMGRMTG 84

Db 71 SVNEGSVLPEDGPIRPFGTIADIQDLPKAKGSGQPIPEALFWLLLTGVPPTAOQVNL 130

Qy 85 LLMETSLDDDEGIRGPIYSIECCQKVLPAKPGGEPLPEGLLWLLLTGKVPSEQVDLS 144

Db 131 SADLMRSRLPSHWOLLNLPDLHPMAQFSIAVTALFESKFKAKVAOGISQDQWSY 190

Qy 145 SQLRSLRAIYPOHYKTIADLPVTAHPTQTPATGVALQVQSEFOKAYEKGHRSKLWEP 204

Db 191 TFEISDLGLKLPVIAAKYIRNVFKDGKMGVEYDPNADYAKNLVNLIGSKDEDFVDLMRLY 250

```

QY 397 LAKL-NPWPNDVDAHSGVLLNYGLTEARYYVLFVGSRALGICSLQIWDRLGLPLERPK 455
Db 456 SYSTDKWIEI-CKK 468
QY 456 SVTME-WLENHCK 468

RESULT 9
ENTRY 14-nm filament protein/citrate synthase (EC 4.1.3.-) -
TITLE Tetrahymena thermophila (SGC5)
ALTERNATE_NAMES 49k filament-forming protein
ORGANISM #formal name Tetrahymena thermophila
DATE 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change
18-Sep-1998
ACCESSIONS JC5625
REFERENCE JC5625; JN0130
#authors Takeda, T.; Watanabe, Y.; Numata, O.
#journal Biochem. Biophys. Res. Commun. (1997) 237:205-210
#title Direct demonstration of the bifunctional property of
Tetrahymena 14-nm filament protein/citrate synthase
following expression of the gene in Escherichia coli.
#accession JC5625
#molecule_type mRNA
#residues 1-462 #label TAK
REFERENCE JN0130
#authors Numata, O.; Takemasa, T.; Takagi, I.; Hirono, M.; Hirano, H.;
Chiba, J.; Watanabe, Y.
#journal Biochem. Biophys. Res. Commun. (1991) 174:1028-1034
#title Tetrahymena 14-nm filament-forming protein has citrate
synthase activity.
#cross-references MUID:91128358
#accession JN0130
#molecule_type mRNA
#residues 1-462 #label NUM
#note ##cross-references GB:D90117; NID:g217406; PID:dl014848; PID:g1688046
part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
COMMENT This protein is involved in oral morphogenesis preceding binary
fission, and in nuclear events during fertilization, such as
formation of gametic pronuclei and zygote formation of gametic
pronuclei. It also acts as a mitochondrial enzyme, citrate
synthase.
GENETICS
#genetic_code SGC5
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; cytoskeleton; mitochondrion;
oxo-acid-lyase
FEATURE
1-21 #domain transit peptide (mitochondrion) #status
predicted #label TNPV
22-462 #product 14-nm filament protein/citrate synthase #status
experimental #label MAT
SUMMARY #length 462 #molecular-weight 52575 #checksum 9423

Query Match 40.2%; Score 1385; DB 2; Length 462;
Best Local Similarity 49.0%; Pred. No. 1,29e-235;
Matches 215; Conservative 89; Mismatches 123; Indels 12; Gaps 10;

Db 25 NLKVVIAELIPKQALKEVREKYGDKV-VQGYTVKQVIGGMRGKMLGSLSRCDPYQG 83
QY 39 DLRSLEQLIPEQDRLLKLLSEHG-KVOLGNITVDVYLGMRGKMTGLLWTSLLDPDEG 97
Db 84 IIFRGYITPOLKEFLPKADPAADQANOEPLPEGIFWLLMTGOLPHTAQVDALKHEWNR 143
QY 98 IFRFGISYECQKLP-A-AKPG--G--EPLPEGLLWLLTGKVPKQVDSLSQELRSP 151
Db 144 GTVQDCVNFILNPKDLHSMTLSMALLYLQKDSFKAKLYDEGKISKDYWEFFEDSM 203
QY 152 ATPDVHVKTIDALPVTAPMTQFATGMALQVQSEFOKAYEKG-IHKSKLWEPTYEDSM 210
Db 204 DLIAKIPRAAIIYRHKYRDSKLIDSDSKLDWAGNYAHMMGFQEHVYKECIRGLSIHCD 263

```

```

QY 211 SLIAQVPLVAAYYRRMYKNGNTIPKDDSLDYCANFAHMLGFSSSDMHMLKLYVTIHS 270
Db 264 HEGGNYSATHTHLVGSALSDPYLSYAGVNGLAGPLHGLANOVLKWLQFIEEGTKVS 323
QY 271 HEGGNYSATHTGLVASALSDPYLSFAAALNLAGPLHGLANOVLWIKSVVEECGENIS 330
Db 324 DKDIEDYVDHVISSGRVWPQYGHAVLRDTPRPHQVDFSKFHLKDDQMIKLHQCADVI 383
QY 331 KEQLKDYAWKTLKSGKVPQFGHGLRKTDPRYTCQREFALKHLPEDPLFQVLAKLYEVF 390
Db 384 PKLLTVKTLANYPVNDVCHSGVLLYSGLTEQYQYTVVFAVSRLG-CMANLIWSRAFG 442
QY 391 LQFLONLAKL-NPWPNDVDAHSGVLLNYGLTEARYYVLFVGSRALGIC-SQLIWDRLG 448
Db 443 LPIERPGSADLKWFHDKYR 461
QY 449 LPLERPKSVTMEWLENHCK 467

RESULT 10
ENTRY citrate (si)-synthase (EC 4.1.3.7) - yeast (Saccharomyces
TITLE cerevisiae)
ALTERNATE_NAMES protein Lp21w; protein YP9723.01; protein YPR001w
ORGANISM #formal name Saccharomyces cerevisiae
DATE 19-May-1995 #sequence_revision 01-Sep-1995 #text_change
06-Feb-1998
ACCESSIONS S52814; S57724; S59746
REFERENCE S52814
#authors Pearson, D.; Bowman, S.
#submission submitted to the EMBL Data Library, April 1995
#accession S52814
#molecule_type DNA
#residues 1-486 #label PE4
##cross-references EMBL:Z48951; NID:g762999; PID:g763000; MIPS:YPR001w
##experimental_source strain AB972
REFERENCE S57724
#authors Jia, Y.K.; Becam, A.M.; Slonimski, P.P.; Herbert, C.J.
#submission submitted to the EMBL Data Library, June 1995
#accession S57724
#molecule_type DNA
#residues 1-486 #label JIA
##cross-references EMBL:X88846; NID:g895888; PID:g895889
REFERENCE S59746
#authors Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.;
Hall, J.; Storms, R.K.; Vo, D.H.; Winnett, E.
#submission submitted to the EMBL Data Library, July 1995
#description The sequence of Saccharomyces cerevisiae chromosome XVI right
arm.
#accession S59746
#molecule_type DNA
#residues 1-486 #label WAN
##cross-references EMBL:U31900; NID:gl276597; PID:g939735; MIPS:YPR001w
GENETICS
#gene SGD:CIT3
#map_position 16R
#cross-references SGD:S0006205; MIPS:YPR001w
KEYWORDS carbon-carbon lyase; oxo-acid-lyase
SUMMARY #length 486 #molecular-weight 53811 #checksum 9998

Query Match 35.4%; Score 1218; DB 2; Length 486;
Best Local Similarity 45.0%; Pred. No. 6.61e-203;
Matches 205; Conservative 102; Mismatches 124; Indels 25; Gaps 15;

Db 21 IIKSSALTLEALENVIPKRDVAKKLCACGYSTFGVPTISSVLGMRGNSMFWQGIS 80
QY 32 VOTSSGLDURSEQLIPEQDRLLKLLSEHGKVLQGNITVDVYLGMRGKMTGLLWTSLS 91
Db 81 LDPEHGKIFQGLTIEBCQNRLPNTGIDGNFSPESMLWLLMTGGVTFQQAASFKKELAI 140
QY 92 LDPDEGIRFGLSIYECQKVLPAKPGEP-LPEGLLWLLTGKVPKQVDSLSQELRS 150

```

```

Db 141 PGRKPLPHYTEKVLSSLPKDPHMTOLAIGLASNMKGSFATNYQKLGKMEFWKDTLED 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 PA-TVDPHYVKIDALPVTAHPMTQFATGVMALQVSEFQKAYXG-IHKSILWEPTVED 208
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 SUNLIASLPLLRIGRYSNTNTEGHPGQVSEEDVCTNCSLLGWTGNTSNTCNLTQS 260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 SMSLIAQVPLVAAYVRRMYKNGNTIPK-DDSLDYCANFAHMLGFS---SSD---M--- 257
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 OSUDEINMLRLTGTIHVDHEGNGNSAHTTHLVGSALSODPYSLSYSSGIMGLAGPLHGLAAQ 320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 H-E---LMLKLYTHSDHEGNGNSAHTGHLVASALSDPYSLSFAALNGLAGPLHGLAAQ 312
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 EVVRFLIEMNSNIS-SIAREOEIKDYLMKILNSRNPVPGYHVLKRPDPREFAMLEFAQ 379
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 EVLLWIKSVVEEGENISKEQ-LKDYANKTLKSGKVPVGFHGVLRKTDPRYTQOREFAL 371
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 KPIERENDKNVLLMOKLAEIAPKVLLEHGKSKNFPNVDNSAGILFHYHIGRELLFTV 439
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 KH-LP-E-DPLFQLVAKLYEVFLQFLONLAKL-NPMPNVDHAHSGVLLNYGLTEAPYTV 427
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 IFGCSRAMGPLTOLVMDRLGLPIERPKNLSNLEGLE 475
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 LFGVSRALGICSLIWDRLGLPLERPKSVTMEWLE 463

RESULT 11
ENTRY I40717 #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) - Corynebacterium
ORGANISM glutamicum
DATE #formal_name Corynebacterium glutamicum
16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
17-Jul-1998
ACCESSIONS I40717
REFERENCE I40717
#authors Eikmanns, B.J.; Thum-Schmitz, N.; Eggeling, L.; Luedtke,
K.O.; Sahm, H.
#journal Microbiology (1994) 140:1817-1828
#title Nucleotide sequence, expression and transcriptional analysis
of the Corynebacterium glutamicum gita gene encoding
citrate synthase.
#accession I40717
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-437 #label RES
#cross-references EMBL:X66112; NID:g505580; PID:g505581
GENETICS gita
#gene homohexamer
COMPLEX
FUNCTION
#description carbon-carbon lyase; oxo-acid-lyase; catalyzes synthesis of
citrate from acetyl-CoA and oxaloacetic acid
#pathway tricarboxylic acid cycle
CLASSIFICATION #superfamily citrate (si)-synthase
KEYWORDS carbon-carbon lyase; homohexamer; oxo-acid-lyase;
tricarboxylic acid cycle
FEATURE
316
SUMMARY #active_site His #status predicted
#length 437 #molecular_weight 48929 #checksum 2093

Query Match 10.5%; Score 360; DB 2; Length 437;
Best Local Similarity 30.4%; Pred. No. 2,58e-40;
Matches 96; Conservative 89; Mismatches 105; Indels 26; Gaps 22;

Db 97 LLINGELPTDELHKFNDEIRHH-TLDEDFKSKQFNVPDRDAHPMATLASSYNIIST-Y- 153
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 LLLTGKVPKSEQVDSLSQELRSRATVPDHVYKT-IDALPVTAHPMTQFATGVMALQVQSE 187
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 YQQLN-PLDEAQL-D-K--ATVRLMAKVPMLAAVAHRAR-KCAPMYTPDNSLNARENFL 207
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 FQKAYEKGIHKSILWEPTVEDSLSIAQVPLVAAYVRRMYKNGNTIPKDDSLDYCANFA 247
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 RMFGVPTPEYEDPTMVAKDLKLLIHADHEQ-NCSTSTVWIGSAQNMVSTAGGN 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 248 HML-GFSS-S-DMHELM-K-L--YVTIHSDEGNGNSAHTGHLVASALSODPYSLSFAALN 300
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 ALSPLHGGAQAVLEMLEDIKNHGGD-ATEFMNK-V-KNKEDGVRLMCFGRHVYKNYD 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 GLAGPLHGLANQEVLLWIKSVVEEGENISKEQLKDYANKTLKSGKVPVGFHGVLRKTD 360
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 PRAAIVETAHEILEHUGGDDLLDLAKLEBEIALA--DDVFISKRKLPNVDYFTGLIYRAM 382
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 361 PRYTQOREFA--LKHLPEDPLFQLVAKLYEVFLQFLONLAKLNPMPNVDHAHSGVLLNY 417
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 GFPTD-PFTVLAIGR 397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 GLTEARYTTLFGVSR 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
ENTRY A43936 #type complete
TITLE citrate synthase - Bacillus sp.
ORGANISM #formal_name Bacillus sp.
DATE 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
21-Aug-1998
ACCESSIONS A43936
REFERENCE A43936
#authors Schendel, F.J.; August, P.R.; Anderson, C.R.; Hanson, R.S.;
Flickinger, M.C.
#journal Appl. Environ. Microbiol. (1992) 58:335-345
#title Cloning and nucleotide sequence of the gene coding for
citrate synthase from a thermotolerant Bacillus sp.
#cross-references MUID:92171501
#accession A43936
#status preliminary
#molecule_type nucleic acid; protein
#residues 1-373 #label SCH
#experimental_source strain C4, ATCC 55182
#note sequence inconsistent with the nucleotide translation
#note sequence extracted from NCBI backbone (NCBIP:84047)
CLASSIFICATION #superfamily citrate (si)-synthase
SUMMARY #length 373 #molecular_weight 41970 #checksum 2911

Query Match 10.1%; Score 347; DB 2; Length 373;
Best Local Similarity 30.3%; Pred. No. 4.98e-38;
Matches 96; Conservative 80; Mismatches 117; Indels 24; Gaps 21;

Db 53 LLEGTIPNEAKQHLBEETLKQYDVPDEIIQVLSLLPKTAHPMDALRTQSVL-A-S-F 109
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 LLLTGKVPKSEQVDSLSQELRSRATVPDHVYKTIDALPVTAHPMTQFATGVMALQVQSEF 188
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 DTELLNREHSTNL-KRAYQ--L-L-GKIPNIVANSYHILHSEEPVQPLQD-LYSANFLY 163
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 QKAYEKGIHKSILWEPTVEDSLSIAQVPLVAAYVRRMYKNGNTIPKDDSLDYGANFAH 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 MITGKPTLEPEKIFDRSLVLYSEHELPN-STFTARVIASLTSDLYGALTGAVASLKGH 222
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 ML-GFSSSDHHE-LMKLYVTHSDHEGNGNSAHTGHLVASALSODPYSLSFAALNGLAGPL 306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 HGSANEAVMEMLQDA-QTV-EGF-KHLLHD---KLSKKEKIM-GFGHRVYKMKMDPRAAM 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 HGLANQEVLLWIKSVVEEGENISKEQLKDYANKTLKSGKVPVGFHGV-LRKTDPRYT 365
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 MKE-ALKELSANGDDLLQWCEAGEQIMRE-EK-GLFPNLDDYAAAPVYKLGPIPLYT 332
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 QREFALKHLEPDLFQLVAKLYEVFLQFLONLAKLNPMPNVDHAHSGVLLNYGLTEARY 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 PIFFS-SRTVGLCAHVM 348
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 TVLFSGVSRALGICSLI 442

RESULT 13
ENTRY YK0PC #type complete
TITLE citrate (si)-synthase (EC 4.1.3.7) - Acetobacter aceti
ORGANISM #formal_name Acetobacter aceti
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change

```


ENTRY	TITLE
S74344	#type complete
	citrate synthase gltA - <i>Synechocystis</i> sp. (strain PCC 6803)

Search completed: Fri Oct 22 17:06:03 1999
Job time : 103 secs.

WORLD

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 17:06:20 1999; Maspar time 14.51 Seconds

Tabular output not generated. 913.625 Million cell updates/sec

Title: >US-08-702-718-6

Description: (1-469) from US08702718.pep

Perfect Score: 3443

Sequence: 1 MVEYRGVSLLSKLSRAVQO.....PLRPKSVTMEWLENHCKKA 469

Scoring table: PAM 150

Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot37

i:swissprot

Statistics: Mean 51.766; Variance 94.679; scale 0.547

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2932	85.2	471	1	CISY_CITMA	CITRATE SYNTHASE, MITO
2	2454	71.3	472	1	CISY_ARATH	CITRATE SYNTHASE, MITO
3	2021	58.7	464	1	CISY_PIG	CITRATE SYNTHASE, MITO
4	1944	56.5	433	1	CISY_CHICK	CITRATE SYNTHASE, MITO
5	1820	52.9	458	1	CISY_CARTEL	PROBABLE CITRATE SYNTH
6	1782	51.8	482	1	CISY_SCHPO	PROBABLE CITRATE SYNTH
7	1763	51.2	479	1	CISY_YEAST	CITRATE SYNTHASE, MITO
8	1740	50.5	467	1	CISY_CANTR	CITRATE SYNTHASE, MITO
9	1725	50.1	474	1	CISY_EMENI	CITRATE SYNTHASE, MITO
10	1714	49.8	475	1	CISY_ASPNG	CITRATE SYNTHASE, MITO
11	1679	48.8	460	1	CISY_YEAST	CITRATE SYNTHASE, MITO
12	1680	48.8	460	1	CISY_NEUCR	CITRATE SYNTHASE, MITO
13	1385	40.2	462	1	CISY_TETTH	CITRATE SYNTHASE, MITO
14	1218	35.4	486	1	CISY_YEAST	CITRATE SYNTHASE 3 (EC
15	360	10.5	437	1	CISY_CORGL	CITRATE SYNTHASE (EC 4
16	347	10.1	373	1	CISY_BACCO	CITRATE SYNTHASE (EC 4
17	310	9.0	436	1	CISY_ACEAC	CITRATE SYNTHASE (EC 4
18	305	8.9	372	1	CISW_BACSU	CITRATE SYNTHASE III (
19	306	8.9	397	1	CISY_SIN3	CITRATE SYNTHASE (EC 4
20	287	8.3	433	1	CISY_ACIAN	CITRATE SYNTHASE (EC 4
21	282	8.2	428	1	CISY_PSEAE	CITRATE SYNTHASE (EC 4
22	280	8.1	376	1	CISY_PYRFU	CITRATE SYNTHASE (EC 4
23	277	8.0	366	1	CISY_THIFE	CITRATE SYNTHASE (EC 4

Query Match

85.2%; Score 2932; DB 1; Length 471;

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	471 AA.
ID	CISY_CITMA			
AC	P49298;			
DT	01-FEB-1996 (REL. 33, CREATED)			
DT	01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)			
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)			
DE	CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).			
GN	CIT.			
OS	CITRUS MAXIMA (PUMPELO) (CITRUS GRANDIS).			
OC	EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			
OC	SAPINDALES; RUTACEAE; CITRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV. SIAMESE SWEET 2240;			
RA	CANEL C.;			
RL	THESIS (1994), UNIVERSITY OF CALIFORNIA AT RIVERSIDE, U.S.A.			
CC	-1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2O) +			
CC	OXALOACETATE.			
CC	-1- PATHWAY: TRICARBOXYLIC ACID CYCLE.			
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.			
CC	-1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE			
CC	METABOLISM.			
CC	-1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: U19481; G624676.			
DR	PROSITE; PS00480; CITRATE SYNTHASE; 1.			
DR	PFAM; PF00285; citrate_synth; 1.			
DR	HSPSP; P23007; 5CSC.			
KW	LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE.			
FT	TRANSIT 1 18			
FT	CHAIN 19 471			
FT	ACT_SITE 307 307			
FT	ACT_SITE 353 353			
FT	ACT_SITE 408 408			
FT	SEQUENCE 471 AA; 52183 MW; A7E3512D CRC32;			

Best Local Similarity 82.3%; Pred. No. 0.00e+00;
Matches 387; Conservative 51; Mismatches 31; Indels 1; Gaps 1;

Db 1 MASLSRATLSRURAGOSNLSNSVRLQMOSSADLDLHLSQLEMPEQOERLKKYKS 60
QY 1 MFVYRGVSLLSKLSRAVQOTNLSNSVRLQVQTSGLDLRSELQELIPEQOERLKKLS 60

Db 61 DLGKAOLGNITIDVIGGMRGTMGLLWETSLDDPDEGIRFGLSIPECKLPAKPGGE 120
QY 61 EHGKVLGNITIDVIGGMRGTMGLLWETSLDDPDEGIRFGLSIPECKLPAKPGGE 120

Db 121 PLPEGLLWLLTGKVPSEKQVDSLSQELSRATVPDHYVKTIDALPVAHPMTQFATGM 180
QY 121 PLPEGLLWLLTGKVPSEKQVDSLSQELSRATVPDHYVKTIDALPVAHPMTQFATGM 180

Db 181 ALQVQSEFOKAYEKGTHKSKWEPTSESLNLIARVPVVAAYVQRIYKDGKLIKFDLSL 240
QY 181 ALQVQSEFOKAYEKGTHKSKWEPTSESLNLIARVPVVAAYVQRIYKDGKLIKFDLSL 240

Db 241 DYGNFSLMGLFDDPKMLRLMYTTHSDHEGNGVSAHTGHLVASALSDPYLSLAALN 300
QY 241 DYGNFSLMGLFDDPKMLRLMYTTHSDHEGNGVSAHTGHLVASALSDPYLSLAALN 300

Db 301 GLAGPLHGLANQEVLLWIKSVDECGENTTEQLDYVWKTLSNGKVVPFGHGVLRKTD 360
QY 301 GLAGPLHGLANQEVLLWIKSVDECGENTTEQLDYVWKTLSNGKVVPFGHGVLRKTD 360

Db 361 PRYTCOREFALKHLPDDPLFOLVSKLYEVVPLTKLGVKNPWPNDVHSGVLLNHGRL 420
QY 361 PRYTCOREFALKHLPDDPLFOLVSKLYEVVPLTKLGVKNPWPNDVHSGVLLNHGRL 420

Db 421 AEARYTVLVFGVSRSLGICSLQIWDRLALPLERPKSVTLDWIKNCKKA 470
QY 421 AEARYTVLVFGVSRSLGICSLQIWDRLALPLERPKSVTLDWIKNCKKA 470

RESULT 2
ID C15Y-ARATH STANDARD; PRT; 472 AA.
AC P20115;
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91370823.
RA UNGER E.A., HAND J.M., CASHMORE A.R., VASCONCELOS A.C.;
RT "Isolation of a cDNA encoding mitochondrial citrate synthase from
Arabidopsis thaliana."
RL PLANT MOL. BIOL. 13:411-418(1989).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X17528; E1188578; -.
CC PIR: JA0149; YKMUM.

DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
DR PRAM; PF00285; citrate_synt; 1.
DR HSP; P23007; SCSC.
KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
KW MULTIGENE FAMILY.
FT TRANSIT 1 ? MITOCHONDRION.
FT CHAIN ? 472 CITRATE SYNTHASE.
FT ACT_SITE 309 309 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 407 407 BY SIMILARITY.
SQ SEQUENCE 472 AA; 52941 MW; 954AFA81 CRC32;

Query Match 71.3%; Score 2454; DB 1; Length 472;
Best Local Similarity 71.7%; Pred. No. 0.00e+00;
Matches 339; Conservative 62; Mismatches 64; Indels 8; Gaps 8;

Db 1 MVFPRSVSAFRLRSRVOGQSSLSNSVRVQMOSSDLDLKSQLOELIPEHKDRLLK 60
QY 1 MVFPRSVSAFRLRSRVOGQSSLSNSVRVQMOSSDLDLKSQLOELIPEHKDRLLK 60

Db 61 SEHGKVLGNITIDVIGGMRGTMGLLWETSLDDPDEGIRFGLSIPECKLPAKPGGE 120
QY 61 SEHGKVLGNITIDVIGGMRGTMGLLWETSLDDPDEGIRFGLSIPECKLPAKPGGE 120

Db 121 GLNHYRFSFVASLNMKCTL-AKSKLKH-CRKTWNRAVSDYVYNAIDALPSTAHPTQFA 178
QY 121 GLNHYRFSFVASLNMKCTL-AKSKLKH-CRKTWNRAVSDYVYNAIDALPSTAHPTQFA 178

Db 179 SGVMAVQSEFOKAYEKGTHKSKWEPTSESLNLIARVPVVAAYVQRIYKDGKLIKFDLSL 240
QY 179 SGVMAVQSEFOKAYEKGTHKSKWEPTSESLNLIARVPVVAAYVQRIYKDGKLIKFDLSL 240

Db 239 DKSLDYGANFSLMGLFDDPKMLRLMYTTHSDHEGNGVSAHTGHLVASALSDPYLSLA 296
QY 239 DKSLDYGANFSLMGLFDDPKMLRLMYTTHSDHEGNGVSAHTGHLVASALSDPYLSLA 296

Db 299 AALNGLAGPLHGLANQEVLLWIKSVDECGENTTEQLDYVWKTLSNGKVVPFGHGVLR 358
QY 299 AALNGLAGPLHGLANQEVLLWIKSVDECGENTTEQLDYVWKTLSNGKVVPFGHGVLR 358

Db 359 RNTDPRYVCOREFALKHLPDDPLFOLVSKLYEVVPLTKLGVKNPWPNDVHSGVLLNH 416
QY 359 RNTDPRYVCOREFALKHLPDDPLFOLVSKLYEVVPLTKLGVKNPWPNDVHSGVLLNH 416

Db 417 YGTEARYTVLVFGVSRSLGICSLQIWDRLALPLERPKSVTLDWIKNCKKA 469
QY 417 YGTEARYTVLVFGVSRSLGICSLQIWDRLALPLERPKSVTLDWIKNCKKA 469

RESULT 3
ID C15Y-PIG STANDARD; PRT; 464 AA.
AC P00889;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN CS.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89000565.
RA EVANS C.T., OWENS D.D., SUMEGI B., KISPAL G., SRERE P.A.;
RT "Isolation, nucleotide sequence, and expression of a cDNA encoding
pig citrate synthase."
RL BIOCHEMISTRY 27:4680-4686(1988).
CC [2]
RP SEQUENCE OF 28-464.
RX MEDLINE; 82231993.
RA BLOXHAM D.P., PARMELEE D.C., KUMAR S., WALSH K.A., TITANI K.;
RT "Complete amino acid sequence of porcine heart citrate synthase."
RL BIOCHEMISTRY 21:2028-2036(1982).

[3]
 RN X-RAY CRYSTALLOGRAPHY (2.7 AND 1.7 ANGSTROMS).
 RX MEDLINE: 83010291.
 RA REMINGTON S., WIEGAND G., HUBER R.;
 RT "Crystallographic refinement and atomic models of two different forms
 of citrate synthase at 2.7- and 1.7-A resolution.";
 RL J. MOL. BIOL. 158:111-152(1982).
 [4]
 RN MUTAGENESIS.
 RX MEDLINE: 91104711.
 RA ALTER G.M., CASAZZA J.P., ZHI W., NEMETH P., SRERE P.A., EVANS C.T.;
 RT "Mutation of essential catalytic residues in pig citrate synthase.";
 RL BIOCHEMISTRY 29:7557-7563(1990).
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 OXALOACETATE.
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATPIX.
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 METABOLISM.
 CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M21197; G164419; -.
 DR PIR; A29966; YKPG.
 DR PDB; 1CTS; 16-JUL-88.
 DR PDB; 2CTS; 09-OCT-88.
 DR PDB; 3CTS; 09-OCT-88.
 DR PDB; 4CTS; 16-JUL-88.
 DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
 DR PFAM; PF00285; citrate_synth; 1.
 KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; 3D-STRUCTURE;
 KW TRANSIT PEPTIDE; METHYLATION.
 FT TRANSIT 1 27 MITOCHONDRION
 FT CHAIN 28 464 CITRATE SYNTHASE.
 FT MOD_RES 395 395 METHYLATION (TRI-).
 FT ACT_SITE 301 301
 FT ACT_SITE 347 347
 FT ACT_SITE 402 402
 FT HELIX 33 55
 FT TURN 56 57
 FT STRAND 59 64
 FT HELIX 65 69
 FT TURN 70 71
 FT STRAND 76 79
 FT STRAND 84 86
 FT TURN 87 89
 FT STRAND 90 92
 FT STRAND 97 97
 FT HELIX 98 104
 FT STRAND 107 107
 FT STRAND 114 114
 FT TURN 116 116
 FT HELIX 117 125
 FT HELIX 131 141
 FT TURN 142 143
 FT HELIX 149 157
 FT TURN 160 161
 FT TURN 164 165
 FT HELIX 166 175
 FT TURN 176 177
 FT HELIX 181 187
 FT TURN 188 188
 FT TURN 192 193
 FT HELIX 194 221
 FT TURN 223 224

FT TURN	232	233
FT TURN	236	237
FT HELIX	238	244
FT TURN	245	245
FT TURN	249	250
FT HELIX	251	261
FT TURN	252	262
FT TURN	266	268
FT HELIX	270	280
FT TURN	281	282
FT HELIX	285	296
FT TURN	297	297
FT TURN	299	309
FT HELIX	310	318
FT TURN	319	319
FT TURN	325	328
FT HELIX	329	337
FT TURN	338	339
FT STRAND	345	345
FT HELIX	355	364
FT TURN	365	367
FT TURN	369	370
FT HELIX	372	375
FT TURN	376	379
FT TURN	381	383
FT HELIX	384	386
FT TURN	387	387
FT HELIX	388	391
FT TURN	392	392
FT STRAND	399	399
FT TURN	402	405
FT HELIX	406	412
FT TURN	413	413
FT HELIX	417	441
FT TURN	442	443
FT STRAND	451	452
FT HELIX	454	461
FT TURN	462	462
SQ SEQUENCE	464 AA; 51629 MW; 3BC3D25 CRC32;	

Query Match 58.7%; Score 2021; DB 1; Length 464;
 Best Local Similarity 61.8%; Pred. No. 0.00e+00;
 Matches 265; Conservative 78; Mismatches 84; Indels 2; Gaps 2;

Db	27	SASSTNLKDILADLIPKEQARIKTRFQGHGNTVVGQITVDMYGGMRGKGLVYETSVD 86
QY	34	TSSGLDLRSELQELIPEQQDRLLKSEHGKVLQGNITVDMVLGGMRGTMGLLWETSLLD 93
Db	87	PDEGIRFGYSIPECOKMLPKAKGGEPLPEGLFWLLVITGQIPTEQVSWLSKEWAKRAA 146
QY	94	PDEGIRFGLSIYECOKVLPAAKPGGEPLPEGLLWLLLTGKVPSEKQVDSLSQELRSRAT 153
Db	147	LPSHVVTMLDNFTNLPMSQLSAAITALNSESFAEYAEIGIRTKYWEIYEDCMDLI 206
QY	154	VPDHVYKTIIDALPVTAPHTQTATGVMAQVQSEFQKAYEKGKHSKLMWEPTYEDMSLI 213
Db	207	AKLPCVAAKIYRNLYREGSSICAIDSKLDMSHNTNMLGYTDAQFTLMRLYLTIHSDHE 266
QY	214	AQVPLVAAYVYRMVYKNGNTI-PKDDSLDYGANFAHMLGFSSSDMHMLKLYVTIHSDHE 272
Db	267	GGNVSATSHLYGSALSDDPYLSFAAAMNGLAGPLHCLANQEVLYWLTQLOKEVGKDVSD 326
QY	273	GGNVSATGHLVASALSDPYLSFAAALNGLAGPLHCLANQEVLYWLTQLOKEVGKDVSD 332
Db	327	KLRDIYINTLNSGRVYPGYGHAVLRKTDPRYTQOREFALKHLPHDPMFKVLAQLYKIVPN 386
QY	333	QLKDYAWKTLKSGKVPYGFHGVLRKTDPRYTQOREFALKHLPHDPMFKVLAQLYKIVPN 392
Db	387	VLEQCKAKNPWPNDVDAHSGVLLQYGYGTEMNYTYTVFVSVRALGVLAQLIWSRALGFLPL 446
QY	393	FLQNLAKL-NPWPNDVDAHSGVLLNYGLTEARYTYTVFVSVRALGICSLQIWDRLGLPL 451
Db	447	ERPKSMSTD 455

```

QY      452 ERPKSVTME 460
|||||:
RESULT      4
ID  C1SV CHICK  STANDARD:      PRT:      433 AA.
AC  P23007:
DT  01-AUG-1991 (REL. 19, CREATED)
DI  01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  CITRATE SYNTHASE, MITOCHONDRIAL (EC 4.1.3.7).
GS.
GN  GALLUS GALLUS (CHICKEN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC  NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN  [1]
RP  X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX  MEDLINE: 90248434.
PA  KAPPUSAS M., BRANCHAUD R., REMINGTON S. J.;
RT  *Proposed mechanism for the condensation reaction of citrate
RT  synthase: 1.9-A structure of the ternary complex with oxaloacetate
RT  and carboxymethyl coenzyme A.*;
RL  BIOCHEMISTRY 29:2213-2219(1990).
RN  [2]
RP  X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF OPEN CONFORMATION.
RX  MEDLINE: 91255228.
PA  LIAO D.-I., KAPPUSAS M., REMINGTON S. J.;
RT  *Crystal structure of an open conformation of citrate synthase from
RT  chicken heart at 2.8-A resolution.*;
RL  BIOCHEMISTRY 30:6031-6036(1991).
CC  -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC  OXALOACETATE.
CC  -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC  -!- SUBUNIT: HOMODIMER.
CC  -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC  -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC  METABOLISM.
CC  -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC  -!- CAUTION: THIS IS AN X-RAY DETERMINED SEQUENCE WHICH WAS
CC  ESTABLISHED USING THE SEQUENCE OF PIG CITRATE SYNTHASE AND
CC  MODIFYING IT BASED ON THE OBSERVED ELECTRON DENSITY.
DR  PDB: 1CSC: 15-APR-91
DR  PDB: 2CSC: 15-APR-91.
DR  PDB: 3CSC: 15-APR-91.
DR  PDB: 4CSC: 15-APR-91.
DR  PDB: 5CSC: 15-APR-91.
DR  PDB: 6CSC: 24-DEC-97.
DR  PDB: 5CTS: 13-OCT-90.
DR  PDB: 6CTS: 15-JAN-93.
DR  PDB: 1CSH: 30-APR-94.
DR  PDB: 1CSI: 30-APR-94.
DR  PDB: 1CSR: 15-OCT-95.
DR  PDB: 1CSS: 15-OCT-95.
DR  PDB: 1AL6: 24-DEC-97.
DR  PDB: 1AM2: 24-DEC-97.
DR  PDB: PS00480: CITRATE_SYNTHASE: 1
DR  PFAM: PF00285; citrate_synth: 1.
KW  LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; 3D-STRUCTURE.
FT  ACT_SITE 274 274
FT  ACT_SITE 320 320
FT  ACT_SITE 375 375
FT  HELIX 6 28
FT  TURN 29 30
FT  HELIX 38 42
FT  TURN 43 47
FT  STRAND 56 59
FT  TURN 60 62
FT  STRAND 63 66
FT  TURN 67 68
FT  STRAND 69 70
FT  HELIX 71 77
FT  STRAND 80 80

```

```

FT TURN 82 83
FT STRAND 87 87
FT HELIX 89 98
FT HELIX 104 117
FT HELIX 122 130
FT TURN 133 134
FT TURN 137 151
FT HELIX 153 160
FT TURN 161 161
FT HELIX 164 194
FT TURN 196 197
FT TURN 205 206
FT TURN 209 217
FT HELIX 218 218
FT TURN 222 234
FT TURN 235 235
FT HELIX 243 253
FT TURN 254 255
FT HELIX 258 269
FT TURN 270 270
FT TURN 272 276
FT HELIX 277 291
FT TURN 292 293
FT HELIX 298 310
FT TURN 311 312
FT TURN 316 317
FT STRAND 318 318
FT HELIX 328 340
FT TURN 342 343
FT HELIX 345 364
FT TURN 365 365
FT STRAND 372 372
FT TURN 374 374
FT HELIX 375 384
FT TURN 385 386
FT HELIX 390 414
FT TURN 415 416
FT HELIX 427 433
SQ SEQUENCE 433 AA; 47373 MW; 6E24FE58 CRC32;

```

```

Query Match 56.5%; Score 1944; DB 1; Length 433;
Best Local Similarity 60.3%; Pred. No. 0.00e+00;
Matches 258; Conservative 73; Mismatches 95; Indels 2; Gaps 2;
Db 1 ASSTNLKVDLAALIPKEQARIKTFQOHGGTALGOITVDMSYGGMKGKGLVYETSVLDP 60
QY 35 SSGDLRSELQELIPEQQDRLLKSEHGKVGOLNITVDMVLGGMRGTGLWETSLDP 94
Db 61 DEGIRFGFSIPCCOKLLPKGXGGEPLPEGLFWLLVTGQIPTGAQVSWLSKEMAKRAAL 120
QY 95 DEGIRFGLSIVECQKVLPAKPGGEPLPEGLLWLLLTGKVPKSEQVDSLSQELRSATV 154
Db 121 PSHVVTMLDNFTNLHPMSQLSAAITALNSESNFARAAYAEGLRTKYWEMYVESAMDLIA 180
QY 155 PDHVKYTIIDALPVTAPMTQFATGVNALQVQSEFOKAYEKGIHKSLEPTYEDSMSLIA 214
Db 181 KLPVAAKIYRNLYRAGSSIGAIIDSKLWSHNFNTNMLGYTDAQFTELMPLYLTIHSDHEG 240
QY 215 QYPLVAAVYVRMYKNGNTI-PKDDSLDYGANFAHMLGFSSDDMHMLKLYVTIHSDEG 273
Db 241 GNVSHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANOEVGLWLAQLOKAXXAGADAS 300
QY 274 GNVSHTSHLVASALSDPYLSFAAAMNGLAGPLHGLANOEVGLWLAQLOKAXXAGADAS 333
Db 301 LRDYIWNLTNSGRVYVPGYGHAVLRKTDPRYTCQREFALKHLPDPMFKLVAOLYKIVPNV 360
QY 334 LKDYAWKTLKSGKVYVPGYGHVLRKTDPRYTCQREFALKHLPDPMFKLVAOLYKIVPNV 392
Db 361 LLEQGAANPNPNVDAHSGVLLQYQYGMTDMNYTTLFGVSRALGLVLAQLINSLRGFFLE 420
QY 393 FLONLAKLPNPNVDAHSGVLLNYGLTEARYTTLFGVSRALGLVLAQLINSLRGFFLE 452
Db 421 RPKSMSTD 428

```

```

|||||:
QY 453 RPKSVTME 460

RESULT 5
ID CISV_CAEEL STANDARD: PRT: 468 AA.
AC P34575;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN T20G5.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SEGERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BERKS M., SMITH A.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z30423; G458482; -
DR PIR: S42370; S42370.
DR WORMPEP: T20G5.2; CE00513.
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synth; 1.
DR HSP: P23007; SCSC.
KW HYPOTHETICAL PROTEIN; LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION;
KW TRANSIT PEPTIDE.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 468 CITRATE SYNTHASE.
FT ACT_SITE 303 303 BY SIMILARITY.
FT ACT_SITE 349 349 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
SQ SEQUENCE 468 AA: 51540 MW; ADA3A630 CRC32;

Query Match 52.9%; Score 1820; DB 1; Length 458;
Best Local Similarity 56.5%; Pred. No. 0.00e+00;
Matches 245; Conservative 82; Mismatches 103; Indels 4; Gaps 4;

Db 25 LSTASGNTLNKLVLSKKIPAHNAKYKSPRTGHTVTVONVINDIYGMRSKMGVMTET 84
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 30 LQVOTSSGLDRSELQELIPEQDRLKSEHGKVGQVGNITVDMVLGMRGTMGLLWET 89
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 85 SVLDPPEGRIFRGYSTPECKLLPKAKGGEPLPEAINWLLCTGDPVSEAOATATKEWN 144
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 90 SLDDPDEGRIFRGLSYIECKVLPAAKPGGEPLPEGLMLLLITGKVPSEQVDSLSQELR 149
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 145 ARADLPTVVRMLDNPDLNHPMAQFIAAALNNESKFPAGAYARGAKASYWEYAYEDS 204
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 150 SRATVPDHYVKTIDALPVTAHPTQFATGMALQVQSEFQKAYEKGIHKSILWEPTYEDS 209
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 205 MDLLAKLPTVAIIYRNLYRDSGVSVIDPKKDWNSANFSSMLGYDDPLFAELMRLYLVIH 264
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 210 MSLIAQVPLVAAYVYRMYKNGNTIP-KDSDLDYGANFAHMLGFSSSOMHELMKLVYTIH 268
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 265 SDHEGGNVSNAHTSHLVGSALSDPYLSFSAAMAGLAGPLHGLANQEVLLNKLIVGEIGN 324
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

|||||:
QY 269 SDHEGGNVSNAHTSHLVGSALSDPYLSFSAAMAGLAGPLHGLANQEVLLNKLIVGEIGN 328
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 325 YTEQLKWTWKHLKSGQVPGYGHAVLRKTDPRYQCOREFALKHLPNDLFLKLVSTLYK 384
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 329 ISKEQLKDYAWKTLKSGKVPVGFHGVGRKTRDPRYTCOREFALKHLPDLFQVLAKLYE 388
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 385 ITPGILLEOGKAKNPWPNVDHSGVLLQYFGMTSEFYTVLFGVSRALG-CLISOLIWARG 443
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 389 VFLOFLONLAKL-NPWPNDVAHSGVLLNYGLTEARYTVVLFVGSRALGIC-SQLWDRA 446
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 444 MGLPLERPKSHSTD 457
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 447 LGLPLERPKSVTME 460
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
ID CISV_SCHPO STANDARD: PRT: 482 AA.
AC Q10306;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROBABLE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN SPAC6C3.03.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69731; E223710; -
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synth; 1.
DR HSP: P23007; SCSC.
KW HYPOTHETICAL PROTEIN; LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION;
KW TRANSIT PEPTIDE.
FT TRANSIT 1 44 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 45 482 CITRATE SYNTHASE.
FT ACT_SITE 319 319 BY SIMILARITY.
FT ACT_SITE 365 365 BY SIMILARITY.
FT ACT_SITE 420 420 BY SIMILARITY.
SQ SEQUENCE 482 AA: 54030 MW; 18145A7D CRC32;

Query Match 51.8%; Score 1782; DB 1; Length 482;
Best Local Similarity 57.2%; Pred. No. 0.00e+00;
Matches 242; Conservative 83; Mismatches 96; Indels 2; Gaps 2.

Db 51 LKDLRLIELPEKQAEIKKFAEHGQDVGIVETINQMGVARGVRSLLIWESVLDPNEGIR 110
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 40 LRSLQLIELPEQDRLKSEHGKVGQVGNITVDMVLGMRGTMGLLWETSLDPEGIR 99
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 111 FRGYTIPCECKLPLSSNGKQPLPESLFWLLVTGEIPTLSQVQALSADWAARSOLPKFVE 170
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 100 FRGLSIYECQKVLPAAPGGEPLPEGLMLLLITGKVPSEQVDSLSQELSRATVPDHY 159
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 171 ELIDRCPTLHPMAOFSLVALEHDSAFKAYERGKMKHDYWKVEYEDCDMLIAKTVP 230
QY 160 KTDALPVTAPHTQATGVMALQVSEFQKAYEKIHKSLWEPTEDMSLIAQVPLV 219
Db 231 AGRIYRLNRYDGVYAPIQMDKHSYNFANVLGFANNEEFVELMRLYLTHADHEGNYSA 290
QY 220 AAVYRRMYKNGNTIPKDDSLDYGANFAHMLGFSDD-MHELMKLYVTIHSDEGNYSA 278
Db 291 HTCHLVGSALSSPFLSMAASNLNGLAGPLHGLANQEVNLFITMKIEGDDLSEETIKSYL 350
QY 279 HTCHLVASALSDDPFLSFAALNGLAGPLHGLANQEVNLFITMKIEGDDLSEETIKSYL 338
Db 351 WKLNSGRVVPVGGHVAFLKRTDPRYTAQREFALEHLEKDPMEQLVSRLEYIIVPGVLTGEG 410
QY 339 WKTLKSGKVPVGGHGVLRKRTDPRYTCQREFALKHLPEDPLFQLVAKLYEVLQFLQNL 398
Db 411 KTKNPYPNVDSHGVLLQYGLKEQSFYTVLFVGSRTLVASQIWDRLALGPIERPKSF 470
QY 399 KL-NPWPNVDAHSGVLLNYGLTEARYTVLVFGVSRALGICSQLIWDRLALGPIERPKSV 457
Db 471 STE 473
QY 458 TME 460

RESULT
ID C1SY YEAST STANDARD; PRT; 479 AA.
AC P00890;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN CIT1 OR LY36 OR GU3 OR YNR001C OR N2019.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=D273-10B;
RA LINDNER P., PLUECKTHIN A.;
RL SUBMITTED (JUL-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D273-10B;
RX MEDLINE; 85003587.
RA SUSSA M., SUDA K., SCHATZ G.;
RT "Isolation of the nuclear yeast genes for citrate synthase and
RT fifteen other mitochondrial proteins by a new screening method.";
RL EMBO J. 3:1773-1781(1984).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S28C / FY1679;
RX MEDLINE; 95208356.
RA VERHASSELT P., AERT R., VOET M., VOLCKAERT G.;
RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
RT the centromere on the Saccharomyces cerevisiae chromosome XIV right
RL arm.";
RL YEAST 10:1355-1361(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF88;
RX MEDLINE; 95028151.
RA LALO D., STETTLER S., MARIOTTE S., GENDREAU E., THURIAUX P.;
RT "Organization of the centromeric region of chromosome XIV in
RT Saccharomyces cerevisiae.";
RL YEAST 10:523-533(1994).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.

```

```

CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z23259; G313750; -
DR EMBL; X00782; G3603; -
DR EMBL; X77395; G496718; -
DR EMBL; Z71616; E239778; -
DR PIR; A01110; YKBY
DR PIR; S35390; S35390.
DR SGD; L0000341; CIT1.
DR PROSITE; PS00480; CITRATE SYNTHASE; 1.
DR PFAM; PF00285; citrate_synth; 1.
DR HSP; P23007; SCSC.
KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
KW MULTIGENE FAMILY.
FT TRANSIT 1 37 MITOCHONDRION.
FT CHAIN 38 479 CITRATE SYNTHASE.
FT ACT_SITE 312 312 BY SIMILARITY.
FT ACT_SITE 358 358 BY SIMILARITY.
FT ACT_SITE 413 413 BY SIMILARITY.
FT CONFLICT 58 58 E -> O (IN REF. 2).
FT CONFLICT 78 78 E -> EE (IN REF. 2).
SQ SEQUENCE 479 AA; 53360 MW; C6896385 CRC32;

Query Match 51.2%; Score 1763; DB 1; Length 479;
Best Local Similarity 55.4%; Pred. No. 0.00e+00;
Matches 237; Conservative 88; Mismatches 102; Indels 1; Gaps 1;

Db 39 SASQTLKEPFAEIIIPAKAEIKFKKEHOKTVIGEVLLQAYGGMRGKIGLVWESVLD 98
QY 34 TSSGLDRLSELQELIPEQDRLKKEHKGVLGNITVDMLVGLMGKMTGLLWETSLLD 93
Db 99 PEEGIRFGRITPEIQRELPAKESGTEPLFALFWLLTGEIPTDAQVKALSADLAARSE 158
QY 94 PDEGIRFGLSIYECQKVLPAKPGGEPLEGLLWLLTGTCKVFSKEQVDSLSQELSRAT 153
Db 159 IPEHVLIQLDLPKDLHPMAQFSIAVTALESSEKFAKAYAGQVSKKEYSYTFEDSLDL 218
QY 154 VPDHVYKTIDALPVTAPHTQATGVMALQVSEFQKAYEKIHKSLWEPTEDMSLI 213
Db 219 GKLPVIAASKYRNVFKDGIITSDPNADYKGNLAQLLGYENKDFIDLMRLYLIIHSDHEG 278
QY 214 AQVPLVAAYVYRRMYKNGNTIPKDDSLDYGANFAHMLGFSDDMHMLKLYVTIHSDEG 273
Db 279 GNVSATHTLVGSALSSPYLSLAAGNLGAGPLHGLANQEVNLFITMKIEGDDLSEETIKSYL 338
QY 274 GNVSATHTLVGSALSSPYLSLAAGNLGAGPLHGLANQEVNLFITMKIEGDDLSEETIKSYL 333
Db 339 IEKLVDTLNLNAGRVVPGYHVAFLKRTDPRYTAQREFALEHLEKDPMEQLVSRLEYIIVPGV 398
QY 334 LKDVAMKTLASGVVPGHGVLRKRTDPRYTCQREFALKHLPEDPLFQLVAKLYEVLQF 393
Db 399 LTKHGKTKNPWPNVDSHGVLLQYGLTEARYTVLVFGVSRALGICSQLIWDRLALGPIE 458
QY 394 LONLAKL-NPWPNVDAHSGVLLNYGLTEARYTVLVFGVSRALGICSQLIWDRLALGPIE 452
Db 459 RPKSFSTE 466
QY 453 RPKSVTME 460

RESULT 8
ID C1SY CANTR STANDARD; PRT; 467 AA.
AC P79024;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

```


Db 342 WSLNAGRVVPGVGHVAVLRKTDPRYTSQREFALRKLPPDPMFKLYSQVYKIPAGVLTG 401
 QY 339 WKTLSGKVVPGVGHVAVLRKTDPRYTSQREFALRKLPPDPMFKLYSQVYKIPAGVLTG 398
 Db 402 KTKNPVNDVHSGVLLQYGLTERNYTVLFGVSRALGVLPOLIIDRAFGAPIERPKSF 461
 QY 399 KL-NPWPNDVHSGVLLQYGLTERNYTVLFGVSRALGVLPOLIIDRAFGAPIERPKSF 457
 Db 462 STE 464
 QY 458 TME 460

RESULT 10
 ID CYSY_ASPPG STANDARD: PRT: 475 AA.
 AC P51044;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
 GN CIT-1.
 OS ASPERGILLUS NIGER.
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
 OC EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WU-2233;
 RA OSHIDA Y., MIYAKE K., KANAYAMA S., KIRIMURA K., USAMI S.;
 RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D63376; G927641; -
 DR PROSITE: P500480; CITRATE_SYNTHASE; 1.
 DR PFAM: PF00285; citrate_synt; 1.
 DR HSP: P23007; 5CSC.
 KW LVASE: TRICARBOXYLIC ACID CYCLE: MITOCHONDRION; TRANSIT PEPTIDE.
 FT TRANSIT 1 ?
 FT CHAIN 2 ?
 FT ACT_SITE 310 475 CITRATE SYNTHASE.
 FT ACT_SITE 336 356 BY SIMILARITY.
 FT ACT_SITE 411 411 BY SIMILARITY.
 FT ACT_SITE 411 411 BY SIMILARITY.
 SQ SEQUENCE 475 AA: 52153 MW: 48D237BC CRC32;
 Query Match 49.8%; Score 1714; DB 1; Length 475;
 Best Local Similarity 57.8%; Pred. No. 0.00e+00;
 Matches 246; Conservative 80; Mismatches 99; Indels 6; Gaps 6;
 Db 36 TGKAKSLKTEFAKPALEKVKRKEGSKV-IGEVLTDOAYGARGVKLVWEGSVL 94
 QY 34 TSSGLDPSLQELIPEQDPLKLLKSEHG-KVQLGNITVDVNLVGMGRGTMGLWETSIL 92
 Db 95 DSEGRIFRGRITPEQELLPRP-PGQPPPLPGLFWLLTGEITEQVDRDLSAEWAR 153
 QY 93 DPDEGRIFRGLSIEYQKYLPAKPGG-EPLPEGLLWLLLTGKVPSEQVDSLSQELSPR 151
 Db 154 SDLPKPIELIDPCSTLHPMSQFSQVLAVALTALEHSAFAKAYAKGINKDYNNYTFEDSD 213
 QY 152 ATPVDHYKIDALPVAHPMTQFATGVNALQVQSEFQKAYEKGIHKSKEFTEDSDS 211

Db 214 LIAKPTIAKIYRNVPKDGKVPAPQKDKDYSYLNALQGYGNNDFVLMRLYLTIHSD 273
 QY 212 LIAQVPLVAVVYRMYKNGNTIPKDDSLDYGANFAHMLGFS-SSDHMLKLVVTHSD 270
 Db 274 HEGGNVAHHTLVGSLSSPMLSLAAGNLGAGPLHGLANQEVLLWLTTKMKAAGNDLS 333
 QY 271 HEGGNVAHHTLVGSLSSPMLSLAAGNLGAGPLHGLANQEVLLWLTTKMKAAGNDLS 330
 Db 334 DEAIKNTLWSTNAGVPGVGHVAVLRKTDPRYTSQREFALRKLPPDPMFKLYSQVYKIA 393
 QY 331 KEQLKDYAWKTLKSGKVVPGVGHVAVLRKTDPRYTSQREFALRKLPPDPMFKLYSQVYK 390
 Db 394 PGVTEHGKTKNPNVNDVHSGVLLQYGLTERNYTVLFGVSRALGVLPOLIIDRALGA 453
 QY 391 LQFLQNLAKL-NPWPNDVHSGVLLQYGLTERNYTVLFGVSRALGVLPOLIIDRALGL 449
 Db 454 PIERPKSYSTE 464
 QY 450 PLERPKSVTME 460

RESULT 11
 ID CYSY_YEAST STANDARD: PPT: 460 AA
 AC P08679;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE, PEROXISOMAL (EC 4.1.3.7).
 GN CIT2 OR YCR05C OR YCR5C OR YCR043.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92254505.
 RA BITEAU N., FREMAUX C., HEBRARD S., MENARA A., AIGLE M., CROUZET M.;
 RT "The complete sequence of a 10.8 kb fragment to the right of the
 RT chromosome III centromere of *Saccharomyces cerevisiae*.";
 RL YCAST 8:61-70(1992).
 CC [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87089811.
 RA ROSENKRANTZ M., ALAM T., KIM K.-S., CLARK B.J., SRERE P.A.,
 RA GUARENTE L.P.;
 RT "Mitochondrial and nonmitochondrial citrate synthases in
 RT *Saccharomyces cerevisiae* are encoded by distinct homologous genes.";
 RL MOL. CELL. BIOL. 6:4509-4515(1986).
 RN [3]
 RP SEQUENCE OF 1-24 FROM N.A.
 RX MEDLINE; 91094853.
 RA LIAO X., SMALL W.C., SRERE P.A., BUTOW R.A.;
 RT "Intramitochondrial functions regulate nonmitochondrial citrate
 RT synthase (CIT2) expression in *Saccharomyces cerevisiae*.";
 RL MOL. CELL. BIOL. 11:38-46(1991).
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z11113; G3299; -
 DR EMBL: M14586; G17127; -

```

DR EMBL; M54982; G171229; -.
DR EMBL; X59720; E264468; -.
DR PIR; A25393; YKBC.
DR YEPD; 5440; -.
DR SGD; L0000342; CIT2.
DR PROSITE; P500342; MICRORODIES_CTER; 1.
DR PROSITE; P500480; CITRATE_SYNTHASE; 1.
DR PFAM; PF00285; citrate_synt; 1.
DR HSP; P23007; 5CSC.
DR LYASE; TRICARBOXYLIC ACID CYCLE; PEROXISOME; MULTIGENE FAMILY.
FT ACT_SITE 293 293 BY SIMILARITY.
FT ACT_SITE 339 339 BY SIMILARITY.
FT ACT_SITE 394 394 BY SIMILARITY.
FT SITE 458 460 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 460 AA; 51413 MW; 9E93E2FD CRC32;

Query Match 48.8%; Score 1679; DB 1; Length 460;
Best Local Similarity 52.4%; Pred. No. 0.00e+00;
Matches 229; Conservative 92; Mismatches 115; Indels 1; Gaps 1;

Db 11 NVASYLQSSQEKTLKERFSEIYTHAODVROFVKEHGKTKISDVLEOVYGMGRGIPG 70
QY 25 NSVNWLOVOTSSGLDRLSELQELIPEQDRLKLLKSEHGKVQGLNITVDMVLGMRGWTG 84
Db 71 SVMGSLVDPEDGIRFGRRTIADIQKLPKAKSSQPLPEALFWLLLTGVEPTQAQVENL 130
QY 85 LLWETSLDDPDEGIRFGLSIVECQKVLPAKPGGEPLEGLLWLLLTGKVPKQVDSL 144
Db 131 SADLMSRSLPSHVOLLNLPKDLHPMAQFSIAVTALESSEKFAKAYAGISKQDYWSY 190
QY 145 SQLSRATVPDHHVYTIIDALPVTAPHTMQFATGVALQVSEFQKAYEKGIHKSILWEP 204
Db 191 TFEDSLDDLGKLPVIAAKIYNNVFKDKMGVEPDNADYAKNLVNLIGSKDEFDVLMRLY 250
QY 205 TYEDSLSIAQPLVAAYVYRMVYKNGNTIPKDDSLDYGANFAHMLGSSSDMHMLKLY 264
Db 251 LTHSHDEGNVSAHTSHLVGSALSPYLSLUSGLNGLAGPLHGRANQEVLEWLFAKKE 310
QY 265 VTHSHDEGNVSAHTSHLVGSALSPYLSLUSGLNGLAGPLHGRANQEVLEWLWIKSV 324
Db 311 VNDYSKDTIEKVLWDLTNSGRVPGYGHAVLRKTDPRYMAQRKFAMDPDYELFKLYS 370
QY 325 CGENISKEQKDYAWKILSKGVVPGFGHGVLRKTDPRYTCQREFALKHLPDPPLFQLVA 384
Db 371 SIYEAPGVLTGKTKNPWPNVDAHSGVLLQYIGLKESFYTVLFGVSRAGFILAQLIT 430
QY 385 KLYEVLFLQLNLAKE-LNPWPNVDAHSGVLLNYGLTEARYTVTLFGVSRALGICSLIW 443
Db 431 DRAIGASIERPKSYSTE 447
QY 444 DRAIGLPLERPKSVTME 460

RESULT 12
ID C1SY_NEUCR STANDARD; PRT; 469 AA.
AC P34085;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN CIT-1
OS NEUROSPORA CRASSA.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC SORDARIALES; SORDARIAACEAE; NEUROSPORA.
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=74A;
RX MEDLINE; 94104594.
RA FEREA T., CONTRERAS E.T., OUNG T., BOWMAN E.J., BOWMAN B.J.;
RT "Characterization of the cit-1 gene from Neurospora crassa encoding
RL the mitochondrial form of citrate synthase.";
RL MOL. GEN. GENET. 242:105-110(1994).
CC -/- CATALYTIC ACTIVITY: CITRATE + COA -> ACETYL-COA + H(2)O +

```

```

QY 331 KEOLXDYAWKTLKSGKVVPGFCHGVRLKTRDPRYTCOREFALKHLPEDPLFQLVAKLYEVF 390
Db 384 PKKLTYKIANPYNVDCHSGVLLYSGLTEYYTYVTFVAVSRALG-CMANLIWSRAFG 442
QY 391 LOFLQNLAKL-NPWNVDASHSVLLNYYGLTEARYTYVLFVGSRALGIC-SOLIWDRAIG 448
Db 443 LPIERPGSADLKWFHDKYR 461
QY 449 LPLERPCKVTMEWLENHCK 467
RESULT 14
ID CDSX_YEAST STANDARD; PRT; 486 AA.
AC P43635;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DT CITRATE SYNTHASE 3 (EC 4.1.3.7).
GN CIT3 OR YPR001W OR YP9723.01.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CW04;
RA JIA Y.K., BECAM A.-M., SLONIMSKI P.P., HERBERT C.J.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA PEARSON D., POWMAN S., BARPEL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WANG Y., AHMED A., BUSSEY H., FORTIN N., FRIESEN J.D., HALL J.,
RA STORMS R.K., VO D.H., WINNETT E.;
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X88846; G895889; -
DR EMBL; 248951; G763000; -
DR EMBL; 271255; E236811; -
DR EMBL; 031900; G939735; -
DR SGD; L0002855; CIT3.
DR PROSITE; PS00342; MICROBODIES_CTER; 1.
DR PFAM; PF00285; CITRATE_SYNTHASE; 1.
DR HSP; P23007; 5CSC.
KW LIASE; TRICARBOXYLIC ACID CYCLE; MULTIGENE FAMILY.
FT ACT_SITE 315 315 BY SIMILARITY.
FT ACT_SITE 361 361 BY SIMILARITY.
FT ACT_SITE 419 419 BY SIMILARITY.
FT SITE 484 486 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 486 AA; 53811 MW; 806F987D CRC32;
Query Match 35.4%; Score 1218; DB 1, Length 486;
Best Local Similarity 45.0%; Pred. No.2.04e-236;
Matches 203; Conservative 102; Mismatches 124; Indels 25; Gaps 15;
Db 21 LIKSSALTIKALENVIPKPKRDVAKKLKACGYSTFGPTITSSVLGGNMGNSMFWOGTS 80

```




W P E R E L E

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Oct 22 17:07:42 1999; MasPar time 29.55 Seconds
Tabular output not generated. 866.114 Million cell updates/sec

Title: >US-08-702-718-6
Description: (1-469) from US08702718.pep
Perfect Score: 3443
Sequence: 1 MFVYRGVSLLSKLSRAVQO.....PLRPKSVTMEWLENHCKKA 469

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: Sptrembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 49.645; Variance 97.075; scale 0.511

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3443	100.0	469	10	024135 CITRATE SYNTHASE (EC 4	0.00e+00
2	3098	90.0	471	10	043175 ETHANOLAMINE AMMONIA-L	0.00e+00
3	3096	89.9	472	10	080433 CITRATE SYNTHASE (EC 4	0.00e+00
4	2987	86.8	474	10	064869 CITRATE SYNTHASE (EC 4	0.00e+00
5	2800	81.3	437	10	096544 CITRATE (SI)-SYNTHASE	0.00e+00
6	2131	61.9	339	10	024259 CITRATE (SI)-SYNTHASE	0.00e+00
7	1997	58.0	466	4	075390 CITRATE SYNTHASE (EC 4	0.00e+00
8	272	7.9	372	1	033929 CITRATE SYNTHASE (CITZ	7.32e-28
9	264	7.7	379	1	032705 CITRATE SYNTHASE (EC 4	2.12e-26
10	260	7.6	379	2	034002 CITRATE SYNTHASE.	1.14e-25
11	253	7.3	411	2	053115 CITRATE SYNTHASE (EC 4	2.11e-24
12	251	7.3	411	2	059777 CITRATE SYNTHASE (EC 4	4.83e-24
13	250	7.3	411	2	077936 CITRATE SYNTHASE (EC 4	7.32e-24
14	252	7.3	431	2	033066 CITRATE SYNTHASE (EC 4	3.19e-24
15	247	7.2	411	2	060047 CITRATE SYNTHASE (EC 4	2.54e-23
16	246	7.1	411	2	059469 CITRATE SYNTHASE (EC 4	3.83e-23
17	245	7.1	411	2	059779 CITRATE SYNTHASE (EC 4	5.80e-23
18	243	7.1	411	2	059775 CITRATE SYNTHASE (EC 4	1.32e-22
19	241	7.0	410	2	059769 CITRATE SYNTHASE (EC 4	3.01e-22
20	238	6.9	356	2	054382 CITRATE SYNTHASE (EC 4	1.03e-21

21 232 6.7 427 2 068883 CITRATE SYNTHASE (EC 4 1.20e+20
22 231 6.7 434 2 094325 CITRATE SYNTHASE (EC 4 1.80e+20
23 214 6.2 429 2 033915 CITRATE SYNTHASE (EC 4 1.70e+17
24 194 5.6 322 2 059236 CITRATE SYNTHASE (EC 4 4.42e+14
25 186 5.4 319 2 059234 CITRATE SYNTHASE (EC 4 9.61e+13
26 181 5.3 315 2 059235 CITRATE SYNTHASE (EC 4 6.43e+12
27 181 5.3 318 2 059238 CITRATE SYNTHASE (EC 4 6.43e+12
28 184 5.3 349 2 030966 CITRATE SYNTHASE (EC 4 2.06e+12
29 184 5.3 390 2 070008 CITRATE SYNTHASE (EC 4 9.39e+12
30 180 5.2 393 2 008395 CITRATE SYNTHASE (EC 4 2.91e+11
31 177 5.1 321 2 059237 CITRATE SYNTHASE (EC 4 2.73e+10
32 171 5.0 319 2 077822 CITRATE SYNTHASE (EC 4 1.81e+03
33 126 3.7 1620 4 099996 YOTIAO (FRAGMENT). 1.81e+03
34 126 3.7 1642 4 014869 CITRATE SYNTHASE (FRAG 1.42e+01
35 112 3.3 299 2 045540 GOLGI COMPLEX AUTOANTI 2.57e+01
36 110 3.2 767 4 092805 P52C6.11 PROTEIN. 6.19e+01
37 107 3.1 283 5 044818 KIAA0378 (FRAGMENT). 8.26e+01
38 106 3.1 808 4 015083 SIMILAR TO BACILLUS SU 1.94e+00
39 102 3.0 188 14 098468 AMINOEXOXYCHORISMATE L 1.46e+00
40 103 3.0 329 2 025309 CELL DIVISION PROTEIN 1.94e+00
41 104 3.0 402 2 085717 REVERSE TRANSCRIPTASE. 2.57e+00
42 103 3.0 480 2 053751 MCASK-A. 1.10e+00
43 102 3.0 664 11 070588 CASK. 2.57e+00
44 102 3.0 897 4 043215 CODED FOR BY C. ELEGAN 1.10e+00
45 105 3.0 2288 5 023081

ALIGNMENTS

RESULT 1 PRELIMINARY: PRT; 469 AA.
ID 024135
AC 024135;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN CITI.

OS NICOTIANA TABACUM (COMMON TOBACCO).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; NICOTIANA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SNN;
RA LA COGNATA U., LANDSCHUETZ V., WILLMITZER L., MUELLER-ROEBER B.;

RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.

CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.

DR EMBL: X84226; E137432;
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LIASE: TRICARBOXYLIC ACID CYCLE.

SQ SEQUENCE 469 AA; 52492 MW; 6BE3C325 CRC32;
Query Match 100.0%; Score 3443; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Db 1 MFVYRGVSLLSKLSRAVQO...PLRPKSVTMEWLENHCKKA 469

QY 1 MFVYRGVSLLSKLSRAVQO...PLRPKSVTMEWLENHCKKA 469

Db 61 EHKVQLGNITVDWV...LGGVGMGTGLLWETSLDPPDEGIFRGLSYECQKVLPAAPGGE 120

QY 61 EHKVQLGNITVDWV...LGGVGMGTGLLWETSLDPPDEGIFRGLSYECQKVLPAAPGGE 120

Db 121 PLPGLLWLLLTGKVP...SKEQVDSLSQELSRATVPDHHVKTIDALPVTAPMTQFATGVM 180

QY 121 PLPGLLWLLLTGKVP...SKEQVDSLSQELSRATVPDHHVKTIDALPVTAPMTQFATGVM 180

```

Db 181 ALQVSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDDSL 240
Qy 181 ALQVSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDDSL 240
Db 241 DYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAALN 300
Qy 241 DYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAALN 300
Db 301 GLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLRKTD 360
Qy 301 GLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLRKTD 360
Db 361 PRYTQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYYGLT 420
Qy 361 PRYTQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYYGLT 420
Db 421 EARYTYTVLFGVSRAIGCSQIWDRLGALPLERPKSVTMEWLENHCKKA 469
Qy 421 EARYTYTVLFGVSRAIGCSQIWDRLGALPLERPKSVTMEWLENHCKKA 469

RESULT 2
ID Q43175 PRELIMINARY: PRT: 471 AA.
AC Q43175;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE ETHANOLAMINE AMMONIA-LYSATE (EC 4.3.1.7).
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. DESIRÉE;
RX MEDLINE: 96073431.
PA LANSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RT "Mitochondrial citrate synthase from potato: predominant expression
in mature leaves and young flower buds.";
RL PLANTA 196:756-764(1995).
CC -!- CATALYTIC ACTIVITY: ETHANOLAMINE - ACETALDEHYDE + NH(3).
CC -!- COFACTOR: COBALAMIN.
DR EMBL: X75082; G483510; -.
DR PFAM: PF00285; citrate_synt.1.
DR MENDEL: 15200; Soltu:2977;mn15200.
KW LYASE.
SQ SEQUENCE 471 AA; 52612 MW; BC96A65C CRC32;

Query Match 90.0%; Score 3098; DB 10; Length 471;
Best Local Similarity 89.6%; Pred. No. 0.00e+00;
Matches 423; Conservative 23; Mismatches 22; Indels 4; Gaps 4;

Db 1 MVFPRSVSLKLRRAVQOSNVSNVWLVQVTSGLDPLSELVQELIPQDDPLKKIK 60
Qy 1 MVFPRSVSLKLRRAVQOSNVSNVWLVQVTSGLDPLSELVQELIPQDDPLKKIK 59
Db 61 SDM-KGSGINITVDMVLGGMGMGTGLLWETSLDDEGIRFRGLSIPECCQVLPAAKPGG 119
Qy 60 SEHGKVLQGNITVDMVLGGMGMGTGLLWETSLDDEGIRFRGLSIPECCQVLPAAKPGG 119
Db 120 EPLPEGLLWLLTGKVPKSEQVNSIVSGIAGSIIIMTTTIDALPVTAPHTQATG 179
Qy 120 EPLPEGLLWLLTGKVPKSEQVNSIVSGIAGSIIIMTTTIDALPVTAPHTQATG 178
Db 180 VMAQVQSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDE 239
Qy 179 VMAQVQSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDD 238
Db 240 SLDYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAA 299
Qy 239 SLDYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAA 297

```

```

Db 300 ALNGLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLR 359
Qy 298 ALNGLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLR 357
Db 360 KTVPRYTQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYY 419
Qy 358 KTDPRYTQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYY 417
Db 420 GLTARYTYTVLFGVSRAIGCSQIWDRLGALPLERPKSVTMEWLENHCKKA 471
Qy 418 GLTARYTYTVLFGVSRAIGCSQIWDRLGALPLERPKSVTMEWLENHCKKA 469

RESULT 3
ID O80433 PRELIMINARY: PRT: 472 AA.
AC O80433;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN DCCS.
OS DAUCUS CAROTA (CARROT).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; ARALIALES; APIACEAE; DAUCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. MS YONSUN;
RA TAKITA E., KOYAMA H., SHIRANO Y., SHIBATA D., HARA T.;
RT "cDNA encoding carrot mitochondrial citrate synthase.";
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
OXIDATIVE METABOLISM.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: AB017159; D1033521; -.
DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 472 AA; 52656 MW; D46C38CC CRC32;

Query Match 89.9%; Score 3096; DB 10; Length 472;
Best Local Similarity 86.2%; Pred. No. 0.00e+00;
Matches 406; Conservative 46; Mismatches 17; Indels 2; Gaps 2;

Db 1 MVFPRSVSLKLRRAVQOSNVSNVWLVQVTSGLDPLSELVQELIPQDDPLKKIK 60
Qy 1 MVFPRSVSLKLRRAVQOSNVSNVWLVQVTSGLDPLSELVQELIPQDDPLKKIK 59
Db 61 AEHGKVLQGNITVDMVLGGMGMGTGLLWETSLDDEGIRFRGLSIPECCQVLPAAKPGG 120
Qy 60 SEHGKVLQGNITVDMVLGGMGMGTGLLWETSLDDEGIRFRGLSIPECCQVLPAAKPGG 119
Db 121 EPLPEGLLWLLTGKVPKSEQVNSIVSGIAGSIIIMTTTIDALPVTAPHTQATG 180
Qy 120 EPLPEGLLWLLTGKVPKSEQVNSIVSGIAGSIIIMTTTIDALPVTAPHTQATG 179
Db 181 VMAQVQSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDD 240
Qy 180 VMAQVQSEFQKAYEKIHKSKLWEPTYESMSLIAQVPLVAAVYRMYKNGNTIPKDD 239
Db 241 LDYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAA 300
Qy 240 LDYGANFAHMLGFSDDHMLKLYVTHSDHEGNGVSAHTGHLVASALSDFYLSFAA 299
Db 301 NGLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLRNT 360
Qy 300 NGLAGPLHGLANOEVLLWIKSVVEECGENISKEQLKDYAMKTLKSGKVPVPGFHVLRNT 359
Db 361 DRYTCQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYY 420
Qy 360 DRYTCQOREFALKHLPEDPLFQVAKLYEVFLQFLONLAKLNPWPNVDASHGVLNYY 418

```



```

Db 421 LTEARYTYVLFVGSRAIGTCSQVWDRALGLPLERPKSVTWMLNHCCKS 471
QY 419 LTEARYTYVLFVGSRAIGTCSQVWDRALGLPLERPKSVTWMLNHCCKA 469

RESULT 4
ID O64869 PRELIMINARY: PRT: 474 AA.
AC O64869;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN F411.16.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA:
OC EUPHYLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS: ROSIDAE;
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
OXIDATIVE METABOLISM.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL; AC094521;
DR PROSITE; PS00480; CITRATE SYNTHASE; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
SQ SEQUENCE 474 AA; 52782 MW; 64F2AC6 CRC32;

Query Match 86.8%; Score 2987; DB 10; Length 474;
Best Local Similarity 82.6%; Pred. No. 0.00e+00;
Matches 389; Conservative 53; Mismatches 27; Indels 2; Gaps 2;

Db 1 MVFFRSVSAFTLRGRVQCSLSNSVRWVWQMSSTDLKLSQELIPEQQDLKKLK 60
QY 1 MVFYRGVSLLSKLSRA-VQTNLSNSVRWLVQVTSGLDLRSELQELIPEQQDLKKLK 59

Db 61 SEHGKVLQGNITVDMVIGMGWMTGLLWETSLLDPEGIRFGLSIPCEQKVLPTAOSGA 120
QY 60 SEHGKVLQGNITVDMVIGMGWMTGLLWETSLLDPEGIRFGLSIPCEQKVLPAKPGG 119

Db 121 EPLPEGLLWLLTGKVPKSEQVEALS KDLANRAAYPDVYVNAIDALPSTAHPTQFASGV 180
QY 120 EPLPEGLLWLLTGKVPKSEQVDSLSQELRSRATVPDHYKITDALPVTAHPTQFATGV 179

Db 181 MALQVQSEFOKAYENGHKSKEWPTYEDCLMLIARVPVAAVYVYRMKNGDSIPSDKS 240
QY 180 MALQVQSEFOKAYENGHKSKEWPTYEDSMSLSIAQVPLVAAVYVYRMKNGTIPKDS 239

Db 241 LDYGANFSHMLGFDEKVKELMRLITITSHDEGNGVSAHTGLVGSALSPLYSAAL 300
QY 240 LDYGANFAHMLGFSSDMHMLKLVYITSHDEGNGVSAHTGLVGSALSPLYSAAL 299

Db 301 NGLAGPLHGLANQVLLWTKSVVEECGEDIISKEQLKEYVWTKLNSGKVPYGGHGVLPNT 360
QY 300 NGLAGPLHGLANQVLLWTKSVVEECGEDIISKEQLKEYVWTKLNSGKVPYGGHGVLPNT 359

Db 361 DRYVQCFEALKHLUPDDPLFOLVSKLYEVVPPVLTGLGKVNPNPNDVDAHSGVLLNHYG 420
QY 360 DRYVQCFEALKHLUPDDPLFOLVSKLYEVVPPVLTGLGKVNPNPNDVDAHSGVLLNHYG 418

Db 421 LTEARYTYVLFVGSRAIGTCSQVWDRALGLPLERPKSVTWMLNHCCKA 471
QY 419 LTEARYTYVLFVGSRAIGTCSQVWDRALGLPLERPKSVTWMLNHCCKA 469

```

```

RESULT 5
ID Q96544 PRELIMINARY: PRT: 437 AA.
AC Q96544;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE) (OXALOACETATE TRANSACETASE)
DE (FRAGMENT).
GN CIT1.
OS BETA VULGARIS (SUGAR BEET).
OC EUKARYOTA: VIRIDIPANTAE: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
OC EUPHYLOPHYTES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS;
OC CARYOPHYLLIDAE; CARYOPHYLLALES; CHENOPODIACEAE; BETA.
RN [1]
RP SEQUENCE FROM N.A.
RA LA COGNATA U. LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER B.;
RL PLANT CELL PHYSIOL. 0:0-0(0).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OXALOACETATE.
CC EMBL; X84228; E137433; -.
DR PFAM; PF00285; citrate_synth; 1.
DR WENDEL; 15201; BETvu:2977;ma15201.
KW LYASE.
FT NON_TER 1 1
SQ SEQUENCE 437 AA; 48881 MW; 003C88CE CRC32;

Query Match 81.3%; Score 2800; DB 10; Length 437;
Best Local Similarity 85.7%; Pred. No. 0.00e+00;
Matches 373; Conservative 34; Mismatches 27; Indels 1; Gaps 1;

Db 1 SSNLDLRSELQELIPEQQDLKKIKKEFGSFLQNGNIVDMVLGMRGTGLLWETSLLDP 60
QY 35 SSGLDLRSELQELIPEQQDLKKIKKEFGSFLQNGNIVDMVLGMRGTGLLWETSLLDP 94

Db 61 EEGIRFGESIPCEQKLLPAASAGAEPLPEGLLWLLTGKVPKSEQVDSLSQELRSRATV 120
QY 95 DEGIRFGELSIYCEQKVLPAKPGGEPLEGLLWLLTGKVPKSEQVDSLSQELRSRATV 154

Db 121 PDHYKITDALPVTAHPTQFCTGVNALQTRSEFQKAYEKHKSKEWPTYEDCLSLIA 180
QY 155 PDHYKITDALPVTAHPTQFATGVNALQVSEFQKAYEKHKSKEWPTYEDSLSLIA 214

Db 181 QVPVAAVYVYRMKNGQVPLDDSLDYGNFAHMLGFDSQPOMLEMLRYITVTHSDEGG 240
QY 215 QVPLVAAVYVYRMKNGTIPKDDSLDYGNFAHMLGFSSDMHMLKLVYITVTHSDEGG 274

Db 241 NVSAHTGLVGSPLSDPYLSFAAALNGLAGPLHGLANQVLLWTKSVVEECGNIETQL 300
QY 275 NVSAHTGLVGSPLSDPYLSFAAALNGLAGPLHGLANQVLLWTKSVVEECGNIETQL 334

Db 301 KDYVWTKLNSGKVPYGGVGLGVRKTDPRYTCQREFALKHLUPDDPPFQVSKLYEVVPPIL 360
QY 335 KDYVWTKLNSGKVPYGGVGLGVRKTDPRYTCQREFALKHLUPDDPPFQVSKLYEVVPPIL 394

Db 361 LEIGKVNPNPNDVDAHSGVLLNHYGITVTHSDEGGVSAHTGLVGSALSPLYSAAL 420
QY 395 ONLAKL-NPNPNVDAHSGVLLNHYGITVTHSDEGGVSAHTGLVGSALSPLYSAAL 453

Db 421 PKSVTWMLNHCCKR 435
QY 454 PKSVTWMLNHCCK 468

RESULT 6
ID O24259 PRELIMINARY: PRT: 339 AA.
AC O24259;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME)
DE (CITRATE CONDENSING ENZYME) (CITROGENASE)
DE (OXALOACETATE TRANSACETASE).

```

GN CIT1.
 OS POPULUS DELTOIDES X POPULUS BALSAMIFERA SUBSP. TRICHOCARPA.
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTES;
 OC EUPHYLLIPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC VIOLETACEAE; SALICACEAE; POPULUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LA COGNATA U. LANDSCHUETZE V., WILLMITZER L., MUELLER-ROEBER R.;
 RL PLANT CELL PHYSIOL. 0:0-0(0).
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 DR EMBL; X84227; E276838; .
 DR PFAM; PF00285; citrate_synth; 1.
 KW LYASE.
 SQ SEQUENCE 339 AA; 37834 MW; B2DA24A9 CRC32;
 Query Match 61.9%; Score 2:31; DB 10; Length 339;
 Best Local Similarity 84.5%; Pred. No. 0.00e+00;
 Matches 284; Conservative 29; Mismatches 22; Indels 1; Gaps 1;
 Db 1 MPTKEQVGSALSKELRDLALVHDYVFKADALPVTAPHTQFATGVMALQVQSEFOKAYEK 60
 QY 135 VPSKEQVDSLSKELSRATVPDHYVTKTIDALPVTAPHTQFATGVMALQVQSEFOKAYEK 194
 Db 61 GHKSKYWEPTVEDSLSIARPIVASIYRIYKDGKVPIMNDSLVIGGNFSHMLGFDS 120
 QY 195 GHKSKLWEPTVEDSMLSLIAQVPLVAAYVRYMYKNGNTIPKDSLDYGANFAHMLGFSS 254
 Db 121 PEMQELMRLVYTNHSDHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEV 180
 QY 255 SDHMLKMLVYTIHSDHEGNGVSAHTGHLVASALSDPYLSFAAALNGLAGPLHGLANQEV 314
 Db 181 LLIWIKSVVEECGENTITTEQLKGVYKNTLSKGVVPGFGHGLRKTVPRTYTCOREFALKHL 240
 QY 315 LLIWIKSVVEECGENTISKQDKYAKNTLSKGVVPGFGHGLRKTVPRTYTCOREFALKHL 374
 Db 241 PDDPLFQVSKLYEYVPPVLTQLGKVKPNPNDVDAHSGVLLNYGLTEARYYTVLFGVSR 300
 QY 375 PEDPLFQVAKLYEYFLOFQNLAKL-NPWPNDVDAHSGVLLNYGLTEARYYTVLFGVSR 433
 Db 301 STIGCSQLIWDALGLPLERPKSVTMELENHCKKA 336
 QY 434 ALGICSQLIWDALGLPLERPKSVTMELENHCKKA 469
 RESULT 7
 ID 075390 PPELIMINARY; PRT; 466 AA.
 AC 075390.
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE (EC 4.1.3.7).
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HEART;
 RA GOLDENTHAL M.J.;
 RL "Cloning and sequence analysis of human citrate synthase cDNA."
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
 CC OXIDATIVE METABOLISM
 CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES
 DR EMBL; AF047042; G3288815; .
 DR PROSITE; PS00480; CITRATE-SYNTHASE; 1.
 KW LYASE; TRICARBOXYLIC ACID CYCLE.
 SQ SEQUENCE 466 AA; 51706 MW; C0532604 CRC32;
 Query Match 58.0%; Score 1997; DB 4; Length 466;
 Best Local Similarity 29.5%; Pred. No. 7.31e-28;

Best Local Similarity 62.2%; Pred. No. 0.00e+00;
 Matches 267; Conservative 76; Mismatches 84; Indels 2; Gaps 2;
 Db 27 SASSTNLKADILADLPKEQARIKTFROQHGKTVVGOITVDMYGGMRKGLVYETSVID 86
 QY 34 TSSGLDLRSELQELIPEQODRLKKLSEHGKVLGNITVDMVLGMRGTMGLLWETSILD 93
 Db 87 PDEGRFRGFSIPCEQKLLPKAKGGEPLPEGLFWLLVTCIPTERQVSWLSKEMAKRAA 146
 QY 94 PDEGRFRGFLSIYECQKVLPAKFGPEPLPEGLLWLLTGKVPSEKQVDSLSQELSRAT 153
 Db 147 LPSHVVTMLDNFPTNLHPMSOLSAAVTALNSFNAGAYARGISKT KYWELIYEDSVDL 206
 QY 154 VPDHYVKTIDALPVTAPHTQFATGVMALQVQSEFOKAYEKHKSILWETPTEDSM 213
 Db 207 AKLPCVAAKIYRNLYWEGSGIGAIIDSLNLSHNFTNNMLGYTDHQTFLMRLYLIHSDHE 266
 QY 214 AQVPLVAAVYVRYMYKNGNTI-PKDDSLDYGANFAHMLGFSSDMHMLKMLVYTIHSDHE 272
 Db 267 GGNYSHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANQEVILWLTQLOKEVCKOVSD 326
 QY 273 GGNYSHTSHLVGSALSDPYLSFAAALNGLAGPLHGLANQEVILLTKSVVEECGENSE 332
 Db 327 KLIRDYIMTILNSGRVPGYGHVLRKTDPRTYTCOREFALKHLPNDPMFKLVAOLYKIVPN 386
 QY 333 OLKQYAWKTLKSGVVPFGHGLRKTDPRTYTCOREFALKHLPEDPLFQVAKLYEYFLO 392
 Db 387 VLEOGKAKNPWPNVDAHSGVLLQYQYOMTMYTYTVLFGVSRALGVLAQLINSGALGFPL 446
 QY 393 FLQNLAKL-NPWPNDVDAHSGVLLNYGLTEARYYTVLFGVSRALGICSQLIWDALGLPL 451
 Db 447 ERPKSMSTE 455
 QY 452 ERPKSVTME 460
 RESULT 8
 ID 028929 PPELIMINARY; PRT; 372 AA.
 AC 028929.
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE (CITZ).
 GN AF1340.
 OS ARCHAEoglobus FULGIDUS.
 OC ARCHAEA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae;
 OC ARCHAEoglobus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE; 98049343.
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY R.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., FEICH C.I., MCNEIL L.K., BADGER J.H., GLADEK A., ZHOU L.,
 RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic,
 RT sulphate-reducing archaeon Archaeoglobus fulgidus."
 RL NATURE 390:364-370(1997).
 DR EMBL; AF001011; G2649234; .
 DR TIGR; AF1340; .
 DR PFAM; PF00285; citrate_synth; 1.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 372 AA; 42270 MW; DB9662BC CRC32;
 Query Match 7.9%; Score 272; DB 1; Length 372;
 Best Local Similarity 29.5%; Pred. No. 7.31e-28;

Matches 76; Conservative 64; Mismatches 95; Indels 23; Gaps 18;

Db 44 EEVAYLLYGEPLPKKYELQDEKIELAERELPQI---I-GL-LT-H-LPPYTHPMVVL 96
 QY 124 EGLWLLTGKVPKQVDSLSQELSRATVPDHVYKTDALPVTAHPTQFATGVMALQ 183
 Db 97 TATSYLSGLDKKIA-VRTRETFNKAQDLIAKPTIVAY-YHRI-RTGRNIIPPALEF 153
 QY 184 VQSEFOKAYEKGIIHKSKEPTIYEDSMSLIAQVPLVAAYVYRMKNG-NIPIKDSLDY 242
 Db 154 AANFLYMLGEEPTKTAERALMDLILHAHEL-NASIFAAARIASTADIYACVVAATG 212
 QY 243 GANFAHML-GFSSDMHE-LMKLYVTIHSDEGNGVSAHTGHLVASALSDPYLSFAAAL 300
 Db 213 TLMGPHLGHGAQVGM---R-MLREVASP--RRAE--EYVKPKTEAGERIMGFCHRYVGM 265
 QY 301 GLAGPLHGLANQEVLLWIKSVVEECENISKEQLDKYANKTLKSKVVPFGHGVLRKT- 359
 Db 266 DPRAELLR-YLAKRLAAE 282
 QY 360 DPTYTCOREFALKHLPED 377

RESULT 9
 ID Q32705 PRELIMINARY; PRT; 379 AA.
 AC Q32705;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
 OS HALOBACTERIUM VOLCANII (HALOFERAX VOLCANII).
 OC ARCHAEA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE; HALOFERAX.
 RN [1]
 RP STRAIN-WFD II;
 RC MADDOCKS D.G.; CONNARIS H., HOUGH D.W., DANSON M.J.;
 RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AJ002075; E1154178; -;
 DR PFAM; PF00285; citrate_synt; 1.
 KW LYASE.
 FT NON_TER 379 379
 SQ SEQUENCE 379 AA; 41848 MW; 480EAC2F CRC32;

Query Match 7.7%; Score 264; DB 1; Length 379;
 Best Local Similarity 31.3%; Pred. No. 2.12e-26;
 Matches 76; Conservative 59; Mismatches 90; Indels 18; Gaps 15;

Db 46 EEVYLLWHGALPTGELDAFSDLAHRDLDGGLVDVARELAEODESPMAALRTLVSAM 105
 QY 124 EGLWLLTGKVPKQVDSLSQELSRATVPDHVYKTDALPVTAH-PMTOFATGVMAL 182
 Db 106 SAYDESA-DPE-DV-TDR--EVNLEKAKRITAKMPSVLA-AYAR-FPRGDDYVEPDESIN 158
 QY 183 VQSEFOKAYEKGIIHKSKEPTIYEDSMSLIAQVPLVAAYVYRMKNGNT-IPKDDSLD 241
 Db 159 HAANFLYMLNGEPNEVLAEFTEDMALVLRADH-GLNASTFSAMVTSSTLSDIYSATSAI 217
 QY 242 YGANFAHML-GFSSSD-MHELMKLYVTIHSDEGNGVSAHTGHLVASALSDPYLSFAAAL 299
 Db 218 GTLSGSLHGGANANVMRLKDV--DD-SDMDPTWVKD-A---LDRGERVAGFVHRVYVK 271
 QY 300 NGLAGPLHGLANQEVLLWIKSVVEECENISKEQLDKYANKTLKSKVVPFGHGVLRKT 359
 Db 272 DPR 274
 QY 360 DPR 362

RESULT 10
 ID Q34002 PRELIMINARY; PRT; 379 AA.
 AC Q34002;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CITRATE SYNTHASE.
 GN C15Y
 OS ANTARCTIC BACTERIUM DS2-3R.
 OC BACTERIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DS2-3R;
 RX MEDLINE; 97454284.
 RA GRIKE U., DANSON M.J., RUSSELL N.J., HOUGH D.W.;
 RT "Sequencing and expression of the gene encoding a cold-active citrate
 synthase from an Antarctic bacterium, strain DS2-3R.";
 RL EUR. J. BIOCHEM. 248:49-57(1997).
 DR EMBL; U85944; G2431774; -;
 DR PFAM; PF00285; citrate_synt; 1.
 SQ SEQUENCE 379 AA; 41832 MW; 47DB0E16 CRC32;

Query Match 7.6%; Score 260; DB 2; Length 379;
 Best Local Similarity 27.8%; Pred. No. 1.14e-25;
 Matches 94; Conservative 88; Mismatches 131; Indels 25; Gaps 22;

Db 47 EQVAYLLWNSLNPNDSELKAFVNFERSHRLDENVKGAIDLSTACHPMDVARTANSVLG 106
 QY 124 EGLWLLTGKVPKQVDSLSQELSRATVPDHVYKTDALPVTAHPTQFATGVMALQ 183
 Db 107 --AN-H-A--RA--ODSSP-EANLEKAMSLATFPVVAYDQRRR-RGEELIEPREP-LDY 156
 QY 184 VQSEFOKAYEKGIIHKSKEPTIYEDSMSLIAQVPLVAAYVYRMKNGNTI-PKDDSLDY 242
 Db 157 SANFLWM-TGEEAAPEVVEAFNVSMILYAEH-SFNASTFTARVITSTLADLHSAVTGAI 214
 QY 243 GANFAHMLGFSSSDMHMLKLY-VT-I-HSDHEGNGVSAHTGHLVASALSDPYLSFAAAL 299
 Db 215 GALKGPLHGGANEAVMHTFEEIGIRKDESDAEATRSKAMWDALAOCKKVMGFGHRYVK 274
 QY 300 NGLAGPLHGLANQEVLLWIKSVVEECENISKEQLDKYANK--TLKSGKVVPFGHGVLR 357
 Db 275 NGDSRVPTMKS-ALDAMIKHYPREMLG-LYNGLEAAAMEE-AKOIK-PNLDPYAGPTYNL 330
 QY 358 KTDPRYTCOREFALKHLPED-PLFOLVAKLYEVFLQELQNLAKLNPWPVNDHAHSGVLLNY 416
 Db 331 MGF-DTEFTPLFAARITGTAHIM-EQVADNALIRP 366
 QY 417 YGLTEARYITVLFQVSRALGICSLIWDRALGUPLERP 454

RESULT 11
 ID Q53115 PRELIMINARY; PRT; 411 AA.
 AC Q53115;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DE CITRATE SYNTHASE (FRAGMENT).
 GN GLTA.
 OS RICKETTSIA SP.
 OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
 RC RICKETTSIACEAE; PICKETTSIAE; RICKETTSIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB BACTERIUM;
 RA RAGULT D., RYDKINA E., ROUX V., ERMEEVA M., BALAYEVA N.;
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U59712; G1390003; -;
 DR PFAM; PF00285; citrate_synt; 1.
 FT NON_TER 411 411
 SQ SEQUENCE 411 AA; 46189 MW; 9CCB3291 CRC32;

Query Match 7.3%; Score 253; DB 2; Length 411;
 Best Local Similarity 27.4%; Pred. No. 2.11e-24;
 Matches 88; Conservative 93; Mismatches 110; Indels 30; Gaps 20;

Db 88 LLIIYGEPLSPSSQYNNFTKKVAVHSLVNERLHYLFQTCSSSHPMALAAVGSF---SAF 144

QY 129 LLLTGKVPKQVDSLSQELRSRATVPDHYKTIIDALPVTAPMTQFATGVMAQVQSEF 188
Db 145 ---YP-DLLNFK--EADYELTAIRMIKTIPTAAMS YK--YSIGOPFIYPDNLDTFENF 196
QY 189 QKAYEGIKHSKLWEPTYE--DSMSLIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANF 246
Db 197 LHMW-FATPCIKYKVNPIIKNALNKIFILHADHEQ--NASTSTVRIAGSSGANPFACVSTG 254
QY 247 AHMLGFS--DMH--E-LMK--LY-VTI-HSDHEGNGVSAHTGHLVASALSDFYLSFAAA 298
Db 255 IASLWPAHGAGANEAVINMLKEIGSSSEYIPKYIAKAKDNKDPFRLMGFGRV--YKNYD 312
QY 299 NGLAGPLHGLANQEVLLWIKSV--VEECGENISKEQLKDYANKTLKSGKVVPFGHGVL 356
Db 313 RAAYLKTCCKEVLKELGOLDNNPQLQIAIELEAIALK--DEYFIERKLYPNVDFYSGIYKA 372
QY 357 RKTDPRTYTCOREF--ALKHLPEPLFQVAKLYEVLQFLONLAKLNPWPVNDVDAHSGVLLN 415
Db 372 AMGIP--SQMFTVLFJAIRTGV 391
QY 416 YGLTEARYYTVLFVGSRALG 436

RESULT 12
ID Q59777 PRELIMINARY; PRT: 411 AA.
AC Q59777
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA SP.
OC BACTERIA: PROTEOBACTERIA: ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HA-91;
RA RAOULT D., RYDKINA E., EREMEVA M., BALAYEVA N.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U59731; G1390011;
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1
FT NON_TER 411
SQ SEQUENCE 411 AA; 46162 MW; 9921533A CRC32;

Query Match 7.3%; Score 251; DB 2; Length 411;
Best Local Similarity 27.2%; Pred. No. 4.83e-24;
Matches 87; Conservative 93; Mismatches 112; Indels 28; Gaps 19;
Db 88 LLIYGELPSQEQYNNFTKQVAHSLVNERLHYLFOTFGSSHPMAIMLAAGVSL---SAF 144
QY 129 LLLTGKVPKQVDSLSQELRSRATVPDHYKTIIDALPVTAPMTQFATGVMAQVQSEF 188
Db 145 ---YP-DLLNFK--EADYELTAIRMIKTIPTAAMS YK--YSIGOPFIYPDNLDTFENF 196
QY 189 QKAYEGIKHSKLWEPTYE--DSMSLIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANF 246
Db 197 LHMW-FATPCIKYKVNPIIKNALNKIFILHADHEQ--NASTSTVRIAGSSGANPFACISGTI 255
QY 247 AHMLGFS--DMH--E-LMK--LY-VTI-HSDHEGNGVSAHTGHLVASALSDFYLSFAAA 299
Db 255 IASLWPAHGAGANEAVINMLKEIGSSSEYIPKYIAKAKDNKDPFRLMGFGRV--YKNYDPR 313
QY 300 NGLAGPLHGLANQEVLLWIKSV--VEEC-GENISKEQLKDYANKTLKSGKVVPFGHGVL 357

Query Match 7.3%; Score 251; DB 2; Length 411;
Best Local Similarity 27.2%; Pred. No. 4.83e-24;
Matches 87; Conservative 93; Mismatches 112; Indels 28; Gaps 19;
Db 88 LLIYGELPSQEQYNNFTKQVAHSLVNERLHYLFOTFGSSHPMAIMLAAGVSL---SAF 144
QY 129 LLLTGKVPKQVDSLSQELRSRATVPDHYKTIIDALPVTAPMTQFATGVMAQVQSEF 188
Db 145 ---YP-DLLNFK--EADYELTAIRMIKTIPTAAMS YK--YSIGOPFIYPDNLDTFENF 196
QY 189 QKAYEGIKHSKLWEPTYE--DSMSLIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANF 246
Db 197 LHMW-FATPCIKYKVNPIIKNALNKIFILHADHEQ--NASTSTVRIAGSSGANPFACISGTI 255
QY 247 AHMLGFS--DMH--E-LMK--LY-VTI-HSDHEGNGVSAHTGHLVASALSDFYLSFAAA 299
Db 255 IASLWPAHGAGANEAVINMLKEIGSSSEYIPKYIAKAKDNKDPFRLMGFGRV--YKNYDPR 313
QY 300 NGLAGPLHGLANQEVLLWIKSV--VEEC-GENISKEQLKDYANKTLKSGKVVPFGHGVL 357

Db 314 AAVLKTCCKEVLKELGOLDNNPQLQIAIELEAIALK--DEYFIERKLYPNVDFYSGIYKA 372
QY 358 KTDPRYTCOREF--ALKHLPEPLFQVAKLYEVLQFLONLAKLNPWPVNDVDAHSGVLLN 416
Db 373 MGIP--SQMFTVLFJAIRTGV 391
QY 417 YGLTEARYYTVLFVGSRALG 436

RESULT 13
ID P77936 PRELIMINARY; PRT: 411 AA.
AC P77936
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE (EC 4.1.3.7) (FRAGMENT).
GN GLTA.
OS RICKETTSIA MONTANA.
OC BACTERIA: PROTEOBACTERIA: ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M 5\6;
RA ROUX V., RYDKINA E., EREMEVA M., RAOULT D.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC OXIDATIVE METABOLISM.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR EMBL: U74756; G1658283;
DR PROSITE: PS00480; CITRATE_SYNTHASE; 1.
DR PFAM: PF00285; citrate_synt; 1.
KW LYASE; TRICARBOXYLIC ACID CYCLE.
FT NON_TER 1
FT NON_TER 411
SQ SEQUENCE 411 AA; 46266 MW; A54EALCF CRC32;

Query Match 7.3%; Score 250; DB 2; Length 411;
Best Local Similarity 27.2%; Pred. No. 7.32e-24;
Matches 87; Conservative 93; Mismatches 112; Indels 28; Gaps 19;
Db 88 LLIYGELPSQEQYNNFTKQVAHSLVNERLHYLFOTFGSSHPMAIMLAAGVSL---SAF 144
QY 129 LLLTGKVPKQVDSLSQELRSRATVPDHYKTIIDALPVTAPMTQFATGVMAQVQSEF 188
Db 145 ---YP-DLLNFK--EADYELTAIRMIKTIPTAAMS YK--YSIGOPFIYPDNLDTFENF 196
QY 189 QKAYEGIKHSKLWEPTYE--DSMSLIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANF 246
Db 197 LHMW-FATPCIKYKVNPIIKNALNKIFILHADHEQ--NASTSTVRIAGSSGANPFACISGTI 255
QY 247 AHMLGFS--DMH--E-LMK--LY-VTI-HSDHEGNGVSAHTGHLVASALSDFYLSFAAA 299
Db 255 IASLWPAHGAGANEAVINMLKEIGSSSEYIPKYIAKAKDNKDPFRLMGFGRV--YKNYDPR 313
QY 300 NGLAGPLHGLANQEVLLWIKSV--VEEC-GENISKEQLKDYANKTLKSGKVVPFGHGVL 357
Db 314 AAVLKTCCKEVLKELGOLDNNPQLQIAIELEAIALK--DEYFIERKLYPNVDFYSGIYKA 372
QY 358 KTDPRYTCOREF--ALKHLPEPLFQVAKLYEVLQFLONLAKLNPWPVNDVDAHSGVLLN 416
Db 373 MGIP--SQMFTVLFJAIRTGV 391
QY 417 YGLTEARYYTVLFVGSRALG 436

Query Match 7.3%; Score 250; DB 2; Length 411;
Best Local Similarity 27.2%; Pred. No. 7.32e-24;
Matches 87; Conservative 93; Mismatches 112; Indels 28; Gaps 19;

Db 88 LLIYGELPSQEQYNNFTKQVAHSLVNERLHYLFOTFGSSHPMAIMLAAGVSL---SAF 144
QY 129 LLLTGKVPKQVDSLSQELRSRATVPDHYKTIIDALPVTAPMTQFATGVMAQVQSEF 188
Db 145 ---YP-DLLNFK--EADYELTAIRMIKTIPTAAMS YK--YSIGOPFIYPDNLDTFENF 196
QY 189 QKAYEGIKHSKLWEPTYE--DSMSLIAQVPLVAAYVYRRMYKNGNT-IPKDDSLDYGANF 246
Db 197 LHMW-FATPCIKYKVNPIIKNALNKIFILHADHEQ--NASTSTVRIAGSSGANPFACISGTI 255
QY 247 AHMLGFS--DMH--E-LMK--LY-VTI-HSDHEGNGVSAHTGHLVASALSDFYLSFAAA 299
Db 255 IASLWPAHGAGANEAVINMLKEIGSSSEYIPKYIAKAKDNKDPFRLMGFGRV--YKNYDPR 313
QY 300 NGLAGPLHGLANQEVLLWIKSV--VEEC-GENISKEQLKDYANKTLKSGKVVPFGHGVL 357
Db 314 AAVLKTCCKEVLKELGOLDNNPQLQIAIELEAIALK--DEYFIERKLYPNVDFYSGIYKA 372
QY 358 KTDPRYTCOREF--ALKHLPEPLFQVAKLYEVLQFLONLAKLNPWPVNDVDAHSGVLLN 416
Db 373 MGIP--SQMFTVLFJAIRTGV 391
QY 417 YGLTEARYYTVLFVGSRALG 436

RESULT 14
ID Q33066 PRELIMINARY; PRT: 431 AA.
AC Q33066
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)



Best Local Similarity 78.4%; Pred. No. 0.00e+00;
Matches 371; Conservative 56; Mismatches 41; Indels 5; Gaps 5;

Db 1 MASLRATSLRSPAGQSNLSVRWQSSADLDLSOL-KEMIPQOQLKAYK 59
QY 1 MVFYRSVLLSKLRSAQSNVSVRWLQVQTSSGLDLSELVQELIPEQDRLKK 60

Db 60 SDLGKQAQGNITIDVIGVGNMGWGLLWETSLLDPDEGIRFGLSIPECOKLLPAKPDG 119
QY 61 SDM-KGSIGNITVDMVLGMRGWTGLLWPKHYLDPEGIRFGLSIPECOKVLPAAKPG 119

Db 120 EPLPEGLLLWLTGKVPKSEQVDSKELDRATVDDV-YKADALPVSAPHTOPASG 278
QY 120 EPLPEGLLLWLTGKVPKSEQVDSKELDRATVDDV-YKADALPVSAPHTOPATG 179

Db 179 VMAQVQSEFQAYEKGIHKSWEPTSEDSNLITARIYVVAAYVQRIYKDKIIPKOD 238
QY 180 VMAQVQSEFQAYEKGIHKSWEPTSEDSNLITARIYVVAAYVQRIYKDKIIPKDE 239

Db 239 SLDYGGNFHMLGDFDPKMLEL-MRLYVYIHSDEHGGVNSAHTGHLVASALSDPFLSFLA 297
QY 240 SLDYGGNFHMLGDFDPKMLEL-MRLYVYIHSDEHGGVNSAHTGHLVASALSDPFLSFLA 299

Db 298 ALNGLAGPLHGLANGLVLLWTKSVYDEGENVITTEQLKDYVWKTLSNGKVYVFGHGVLR 357
QY 300 ALNGLAGPLHGLANGLVLLWTKSVYDEGENVITTEQLKDYVWKTLSNGKVYVFGHGVLR 359

Db 358 KTDPRVTCOREFAMKHLDPDLFQVLSKLYEWPPIILTKLGKVKNPVNDVDAHSGVLLNH 417
QY 360 KTDPRVTCOREFAMKHLDPDLFQVLSKLYEWPPIILTKLGKVKNPVNDVDAHSGVLLNH 418

Db 418 FGLAEARYTYVLFVGSRLGICSLQIWDRLALGLPLERPKSVTMDIEKNCKKA 470
QY 419 FGLAEARYTYVLFVGSRLGICSLQIWDRLALGLPLERPKSVTMDIEKNCKKA 471

RESULT 2
ID C15Y ARATH STANDARD: PRT: 472 AA.
AC P20115;
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
SEQUENCE FROM N.A.
MEDLINE: 91370823
UNGER E.A., HAND J.M., CASHMORE A.R., VASCONCELOS A.C.;
*Isolation of a cDNA encoding mitochondrial citrate synthase from
RT Arabidopsis thaliana.*
RL PLANT MOL. BIOL. 13:411-418(1989).
CC -!- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +
CC OXALOACETATE.
CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
CC METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X17528; E1188578;
CC FIR: JA0149; YKUM.

DR PROSITE: PS00480; CITRATE SYNTHASE; 1.
DR PFM; PF00285; citrate_synt; 1.
DR HSP; P23007; 5SCC.
KW LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION; TRANSIT PEPTIDE;
MULTIGENE FAMILY.
FT TRANSIT 1 7 MITOCHONDRION.
FT CHAIN 7 472 CITRATE SYNTHASE.
FT ACT_SITE 309 309 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 407 407 BY SIMILARITY.
SQ SEQUENCE 472 AA; 52941 MW; 954AFA81 CRC32;

Query Match 66.58; Score 2305; DB 1; Length 472;
Best Local Similarity 68.9%; Pred. No. 0.00e+00;
Matches 328; Conservative 60; Mismatches 76; Indels 12; Gaps 12;

Db 1 MVFYRSYSAFRLSRVQGOSSLSNVRWQSSADLDLSOL-OELIPEKHOKRLAKL 59
QY 1 MVFYRSVLLSKLRSA-QVQSNVSVRWLQVQTSSGLDLSELVQELIPEQDRLKK 59

Db 60 KSEHGKVLQGNITVDMVIGVGNMGWGLLWETSLLDPDEEYFALGDCRLPECOKALLPTAQS 119
QY 60 KSDM-KGSIGNITVDMVLGMRGWTGLLWPKHYLDPEGIRFGLSIPECOK-VLPAAKP 117

Db 120 GGLNHYRRSFVSLNWKGLT-AKSKLKH-CRKTWNRAAVSDYV-YNADALPSTAHPTQ 176
QY 118 GG-FPLPEGLL-WLLTGKVPKSEQVNSVSGIAESGIIISLIIMYTTIDALPVTAHPTQ 175

Db 177 FASGVMAQVQSEFQAYEKGIHKSWEPTSEDSNLITARIYVVAAYVQRIYKDKIIPKOD 236
QY 176 FATGVMAQVQSEFQAYEKGIHKSWEPTSEDSNLITARIYVVAAYVQRIYKDKIIPKOD 235

Db 237 PSDKSLDYGANFMSHMLGDFDPKMLEL-MRLYVYIHSDEHGGVNSAHTGHLVASALSDPFL 295
QY 236 PKDESLDYGANFMSHMLGDFDPKMLEL-MRLYVYIHSDEHGGVNSAHTGHLVASALSDPFL 295

Db 296 SFAAALNGLAGPLHGLANGLVLLWTKSVYDEGENVITTEQLKDYVWKTLSNGKVYVFGH 355
QY 296 SFAAALNGLAGPLHGLANGLVLLWTKSVYDEGENVITTEQLKDYVWKTLSNGKVYVFGH 355

Db 356 GVLRLTDPVRYVQVQFALKHHPDDPLFQ-CCKLMKL-ASCLTELESEEPWPVNDVDAHSGVL 413
QY 356 GVLRLTDPVRYVQVQFALKHHPDDPLFQ-CCKLMKL-ASCLTELESEEPWPVNDVDAHSGVL 415

Db 414 LNHVGLTEARYTYVLFVGSRLGICSLQIWDRLALGLPLERPKSVTMDIEKNCKKA 469
QY 416 LNYVGLTEARYTYVLFVGSRLGICSLQIWDRLALGLPLERPKSVTMDIEKNCKKA 471

RESULT 3
ID C15Y.PIG STANDARD: PRT: 464 AA.
AC P00889;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.7).
GN CS.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89000665.
RA EVANS C.T., OWENS D.D., SUMEGI B., KISPAL G., SRERE P.A.;
RT "Isolation, nucleotide sequence, and expression of a cDNA encoding
RL pig citrate synthase."
RL BIOCHEMISTRY 27:4680-4686(1988).
RN [2]
RP SEQUENCE OF 28-464.
RX MEDLINE: 82231993.
RA BLOXHAM D.P., PARMELEE D.C., KUMAR S., WALSH K.A., TITANI K.;
RT "Complete amino acid sequence of porcine heart citrate synthase."
RL BIOCHEMISTRY 21:2028-2036(1982).





Query Match	29.80	Score 564	DB 27	Length 1378
-------------	-------	-----------	-------	-------------

